PB83-260810

Survey of Statistical Measures of Model Performance and Accuracy for Several Air Quality Models

Systems Applications, Inc., San Rafael, CA

Prepared for

Environmental Protection Agency
Research Triangle Park, NC

25 Jun 82

U.S. Department of Commerce
National Technical Information Service

NTIS

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-450/4-83-001	2.	3. RECIPIENT'S ACCESSION NO. PB8 3 260810
4. TITLE AND SUBTITLE A SURVEY OF STATISTICAL MEASURES OF MODEL PERFORMANCE AND ACCURACY FOR SEVERAL AIR QUALITY MODELS		5. REPORT DATE 6/25/82 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Gary E. Moore, Till E. Stoeckenius, Douglas A. Stewart		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Systems Applications, Inc. 101 Lucas Valley Road San Rafael, California 94903		10. PROGRAM ELEMENT NO. 11. CONTRACT/GRANT NO. 68-01-5845
12. SPONSORING AGENCY NAME AND ADDRESS Office of Air Quality Planning and Standards U.S. Environmental Protection Agency Research Triangle Park, North Carolina 27711		13. TYPE OF REPORT AND PERIOD COVERED 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This study reviews a number of model evaluation studies. The statistical measures of model accuracy and performance presented in these studies have been grouped into sets reflecting the recommendations of the American Meteorological Society's (AMS) Woods Hole Workshop on judging air quality model performance. These statistics have been further aggregated into a generalized set of statements that describe model performance and accuracy and are presented in tables organized according to model type and use. Results indicate only four or five models that have a significant evaluation history; in addition they illustrate several recurring tendencies of past model evaluation studies and ways of avoiding certain deficiencies in model evaluation.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Air Pollution Turbulent Diffusion Mathematical Models Computer Models Atmospheric Models	Dispersion Air Quality Simulation Model Model Validation Model Evaluation	

18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)* UNCLASSIFIED	21. NO. OF PAGES 124
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (9-73)

PB83-260810

United States
Environmental Protection
Agency

Office of Air Quality
Planning and Standards
Research Triangle Park NC 27711

EPA-450/4-83-001
November 1982

Air

EPA

A Survey of Statistical Measures of Model Performance and Accuracy for Several Air Quality Models

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA 22161

EPA-450/4-83-001

A Survey of Statistical Measures of Model Performance and Accuracy for Several Air Quality Models

by
Gary E. Moore,
Till E. Stoeckenius,
and Douglas A. Stewart

Systems Applications, Inc.
101 Lucas Valley Road
San Rafael, California 94903

Publication No. 82345

Contract No. 68-01-5845

EPA Project Officer: Russell Lee

Prepared for

U.S ENVIRONMENTAL PROTECTION AGENCY
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

November 1982

i.a

DISCLAIMER

This report has been reviewed by the Office of Air Quality Planning and Standards, U. S. Environmental Protection Agency, and approved for publication as received from Systems Applications, Inc. Approval does not signify that the contents necessarily reflect the views and policies of the U. S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use. Copies of this report are available from the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.

Preface

This report provides summary statements on the accuracy of specific air quality models. Models considered in this study are (1) those recommended by the Environmental Protection Agency (EPA) for routine use and (2) those submitted to EPA by model developers for consideration. Performance measures recommended by the American Meteorological Society are exclusively used in preparing the statements of model accuracy.

This work was performed in response to requests that the accuracy of models be more thoroughly documented. Such requests were made at several public hearings on air quality models that have been conducted by this office. The report represents the most comprehensive attempt ever undertaken to summarize the performance of a large group of air quality models in precise statistical terms. This information should be of assistance in selecting and using the most appropriate models in air pollution regulatory applications.

The reader should take note that the statements of accuracy are based on information made publicly available by the model developer as part of the documentation for each model or through widely circulated publications. Each developer was specifically requested to supply such information. No attempt was made to conduct an extensive literature search. Our philosophy was that the accessibility of such information is the responsibility of the model developer. Users of models can not be expected to routinely search out all information on models to substantiate the claims of accuracy by developers. Typically, users must rely on information for each model that is readily available. As a result, it was found that some of the models had little, if any, publicly available information suitable for documenting their accuracy. For many models such data existed for one or two studies. Only for a few models were several studies available on which to base a statement of model accuracy.

The reader is further cautioned concerning attempts to judge the relative performance of two or more models based on information presented herein. Few of these models have been evaluated with the same data bases, making such performance comparisons speculative at best. EPA is taking steps to correct this deficiency through performance comparisons for a large number of models with similar applications. Plans for evaluating about 40 models in eight model categories over the next several years have been developed. Performance measures recommended by the American Meteorological Society will be used. The results of these comparative evaluations will be made publicly available as they are completed.

Russell F. Lee
Project Officer

CONTENTS

Disclaimer ii

Figures iv

Tables iv

Glossary of Symbols and Terms. viii

1 INTRODUCTION 1-1

2 DEFINING AGRREGATE STATEMENTS AND STATISTICAL MEASURES WHICH DESCRIBE MODEL PERFORMANCE AND ACCURACY. 2-1

  2.1 Development of Aggregate Statements of Model Performance and Accuracy. 2-1

  2.2 Recasting Model Evaluation Statistics into the AMS Recommended Statistics 2-6

  2.3 Specific Methods Used to Calculate Statistics 2-9

3 RURAL POINT SOURCE MODELS. 3-1

4 URBAN AIR QUALITY MODELS 4-1

5 INDUSTRIAL COMPLEX SOURCE MODELS 5-1

6 REACTIVE CHEMISTRY MODELS. 6-1

7 MOBILE SOURCE MODELS 7-1

8 COMPLEX TERRAIN MODELS 8-1

9 PLUME VISIBILITY MODELS. 9-1

10 LONG-RANGE TRANSPORT MODELS. 10-1

11 SUMMARY 11-1

REFERENCES R-1

Preceding page blank

FIGURES

2-1 A Hierarchical tree of statistical performance measures found in the AMS Workshop Study 2-3

TABLES

2-1 Some Aggregate Statements of Model Performance and Accuracy 2-11

3-1 Summary of the Various Rural Air Quality Models 3-5

3-2 Available Statistical Measures for the Rural Point-Source Models 3-6

3-3 A Descriptive Summary of Evaluation Studies for the Rural Point-Source Models 3-7

3-4 Statistical Measures and Aggregate Performance Statements for TEM8 3-8

3-5 Statistical Measures and Aggregate Performance Statements for CRSTER 3-9

4-1 Summary of the Various Urban Air Quality Models 4-4

4-2 Available Statistical Measures for the Urban Air Quality Models 4-5

4-3 Summary of Evaluation Studies of Urban Air Quality Models 4-7

4-4 Statistical Measures and Aggregate Performance Statements for AQDM 4-8

4-5 Statistical Measures and Aggregate Performance Statements for CDM.. 4-9

4-6 Statistical Measures and Aggregate Performance Statements for RAM.. 4-10

5-1 A Summary of Basic Information about the Various Industrial Complex Source Models........................... 5-4

5-2 Available Statistical Measures for Each of the Industrial Complex Source Models........................... 5-5

5-3 A Descriptive Summary of Evaluation Studies for the Industrial Complex Source Models........................... 5-6

5-4 Statistical Measures and Aggregate Performance Statements for ICS(ST).. 5-7

5-5 Statistical Measures and Aggregate Performance Statements for PASDIM.. 5-8

5-6 Statistical Measures and Aggregate Performance Statements for PAL.. 5-9

5-7 Statistical Measures and Aggregate Performance Statements for BLP.. 5-11

6-1 Summary of Basic Information About the Various Reactive Chemistry Models... 6-6

6-2 Summary of the Availability of Model Evaluation Studies and Their AMS Recommended Statistics for the Reactive Models... 6-7

6-3 A Descriptive Summary of Data Bases Used in the Reactive Model Evaluations................................ 6-8

6-4 Summary of AMS Recommended Statistics for ELSTAR........... 6-9

6-5 A Summary of AMS Recommended Statistics for RPM............ 6-10

6-6 A Summary of AMS Recommended Statistics for the Airshed Model... 6-11

7-1 A Summary of Basic Information About the Various Mobile Models.......... 7-4

7-2 A Summary of the Availability of AMS Statistical Measures from Model Performance Evaluation Studies of Mobile Models.......... 7-5

7-3 A Descriptive Summary of Evaluation Studies for the Mobile Source Models.......... 7-6

7-4 Statistical Measures and Aggregate Statements for HIWAY-1.......... 7-7

7-5 Statistical Measures and Aggregate Statements for HIWAY-2.......... 7-8

7-6 Statistical Measures and Aggregate Statements for CALINE3.......... 7-9

7-7 Statistical Measures and Aggregate Statements for GMLINE.......... 7-10

7-8 Statistical Measures and Aggregate Performance Statements for APRAC-1A.......... 7-11

7-9 Summary of the Non AMS Recommended Statistics Appearing in the Mobile Model Evaluation Studies.......... 7-12

8-1 A Summary of Basic Information About the Various Complex Terrain Models.......... 8-4

8-2 Available AMS Statistical Measures from Model Performance Evaluation Studies of Models Used in Complex Terrain.......... 8-5

8-3 A Descriptive Summary of the Evaluation Studies of the Complex Terrain Models.......... 8-6

8-4 Statistical Measures and Aggregate Statements for SHORTZ.......... 8-7

8-5 Statistical Measures and Aggregate Statements for COMPLEX I.......... 8-8

8-6 Statistical Measures and Aggregate Statements for COMPLEX II.......... 8-9

8-7 Statistical Measures and Aggregate Statements for IMPACT.. 8-10

8-8 Statistical Measures and Aggregate Statements for RTDM.WC... 8-11

8-9 Statistical Measures and Aggregate Statements for 3141/4141... 8-12

8-10 Statistical Measures and Aggregate Performance Statements for COMPTER.................................. 8-13

9-1 Statistical Measures and Aggregate Performance Statements for PLUVUE................................... 9-5

9-2 Statistical Measures and Aggregate Performance Statements for ERT Visibility........................... 9-6

10-1 A Summary of Basic Information about the Various Long-Range Transport Models................................ 10-4

10-2 A Summary of the Availability of AMS Statistical Measures from Model Performance Evaluation Studies of Long-Range Transport Models... 10-5

10-3 A Descriptive Summary of Data Bases Used in the Long-Range Models.. 10-6

10-4 Statistical Measures and Aggregate Statements for RTM...... 10-7

10-5 Statistical Measures and Aggregate Statements for RADM..... 10-9

GLOSSARY OF SYMBOLS AND TERMS

Basic Statistical Measures

N	Sample set size
C_o (x, t)	observed concentrations at location x and time t (short hand designation C_o)
C_p (x, t)	predicted concentrations at location x and time t (short hand designation C_p)
Pearson moment correlation coefficient	$r = \Sigma (C_o - \bar{C}_o) (C_p - \bar{C}_p)/[\Sigma (C_o - \bar{C}_o)^2 \cdot \Sigma (C_p - \bar{C}_p)^2]^{1/2}$
cross-correlation function	$r(\Delta t) = \Sigma [C_o(t + \Delta t) - \bar{C}_o) (C_p (t) - \bar{C}_p)]/[\Sigma (C_o - \bar{C}_o)^2 \cdot \Sigma (C_p - \bar{C}_p)^2]^{1/2}$
residual	$d = C_o(x,t) - C_p(x,t)$
average residual (bias)	$\bar{d} = \frac{1}{N} \Sigma d$
standard deviation of residuals (noise)	$S_d^2 = \frac{1}{N - 1} \Sigma (d - \bar{d})^2$
root-mean-square error (gross error)	$RMSE_d = [\frac{1}{N} \Sigma d^2]^{1/2}$
average absolute error	$\lvert\bar{d}\rvert = \frac{1}{N} \Sigma \lvert d\rvert$

$\vec{D}_o(L,t)$	vector distance from a location, L, (where an observed concentration at time t occurred) to some origin
$\vec{D}_p(L',t)$	vector distance from a location, L′, (where a predicted concentration at time t occurred) to some origin
distance vector difference	$\vec{\Delta} = \vec{D}_o(L,t) - \vec{D}_p(L',t)$
unmatched time difference	$\Delta t = t_o - t_p$
t_o	time when nth highest observation occurs
t_p	time when nth highest prediction occurs
$g(C_o)$	cumulative distribution of observations
$f(C_p)$	cumulative distributions of predictions
Z statistic transformation	$\mu_z = \frac{1}{2} \ln \left(\frac{1+r}{1-r}\right)$ where $S_z = \frac{1}{\sqrt{N-3}}$; and z tends to be normally distributed
t statistic (pooled)	$t_{\nu,\alpha}$ where $\nu = N - 1$, the number of degrees of freedom, $1 - \alpha \equiv$ confidence level
Chi square statistic	$\chi^2_{\nu,\alpha}$
average residual $\overline{d}$ confidence interval	$\pm t_{\nu,.05} \left[\frac{S_o^2}{N} + \frac{S_p^2}{N}\right]^{1/2}$
estimate of standard deviation of residuals confidence interval	$\frac{\nu S_d^2}{\chi^2_{\nu,\,.05}} < \sigma_d^2 < \frac{\nu S_d^2}{\chi^2_{\nu,\,.95}}$

Z statistic confidence interval $\mu_z \pm \frac{1.96}{\sqrt{N - 3}}$ (95% level)

Graphical Display Shorthand

CP	cumulative frequency distribution plots of observed and predicted concentrations
H	histograms of observations or predictions
HR	histograms of observed-predicted differences (residuals)
SP	observed-predicted scatter plots
SPIR	scatter plots of input variables versus residuals
SPOR	scatter plots of observations versus residuals
SPR	spatial isopleth plots of residuals
TS	time series plots of predictions versus observations

1 INTRODUCTION

Recently, the U.S. Environmental Protection Agency (EPA) undertook a program to evaluate the performance and accuracy of many of the air quality simulation models now in use. The American Meteorological Society (AMS) Woods Hole Workshop on judging air quality performance (Fox, 1981), has recommended this type of program and has suggested a number of statistical measures from which statements of model performance and accuracy are to be made. This report deals with two aspects of the EPA's model performance evaluation program:

> Structuring of the recommended AMS statistical measures into a concise, generalizable set of statements (called aggregate statements) that describe a model's performance and accuracy.

> Review and recasting of statistical measures from past evaluation studies to render the statistics representative of the AMS recommended statistics, from which accurate model intercomparisons and aggregate statements of performance and accuracy can be made.

First, the means of relating the various AMS statistical measures to general aspects of model performance and accuracy is explained. A small set of aggregate statements summarizing the principal aspects of a model's forecasting ability is developed. These aggregate statements are formulated so as to have equivalent meaning for models of similar designated usage, e.g., point source models, mobile source models, etc.

The major emphasis of this report is on producing a display of AMS performance statistics culled from past evaluation studies. The reader should be able to use these displays with minimum effort to

> Determine the availability of model performance statistics for a particular model.

> Judge whether model performance and accuracy are sufficient, given a set of modeling objectives and criteria.

> Make fair and meaningful comparisons of performance and accuracy between several similar models.

> Judge the adequacy and usefulness of statistics used to evaluate model performance and accuracy.

To achieve this purpose, the AMS Workshop (Fox, 1981) recommended that statistical measures be selected to form a standard set of statistics to be used in reporting on model accuracy. As many as possible of the non-AMS recommended statistics found in model evaluation studies are converted to the AMS forms to make full use of the information in the evaluation studies.

Section 2 discusses the AMS statistical measures and how these measures are used to define and quantify aggregate statements of model accuracy and performance. The remaining sections of this report deals with specific groups of models. The models are arranged first by intended use and second by physical modeling capability. Some general purpose models appear in several sections depending on how they were operated. The bulk of information about the models is presented in tabular form with accompanying comments that serve to clarify and expand on the table.

In cases where several evaluation studies were performed for a specific model, the results may conflict. This is generally because the model was operated in different configurations. No effort is made in this report to resolve such conflicts (e.g. overprediction versus underprediction), and in any event, such a resolution may not be possible at present. On the other hand, in some cases, two or more models may be very similar (or even identical) to each other. Thus, the evaluation studies for one model may provide some information on the accuracy of another as well, provided reasonable caution is used. The reader should excercise care in interpreting any of these results.

2 DEFINING AGGREGATE STATEMENTS AND STATISTICAL MEASURES WHICH DESCRIBE MODEL PERFORMANCE AND ACCURACY

2.1 DEVELOPMENT OF AGGREGATE STATEMENTS OF MODEL PERFORMANCE AND ACCURACY

Generally, a modeling study starts with the definition of specific objectives. Information from performance evaluation studies is often needed to

> Choose a model to apply to a specific situation

> Determine the usefulness and uses of the modeling results.

A model performance evaluation study helps to determine the latter by asking questions about the differences between observations and model predictions. Specific types of differences are referred to as performance attributes. Some examples of performance attributes are

> Degree of randomness of differences

> Degree of over- or underprediction by the model

> Degree of significance of differences.

The performance attributes of interest to a particular modeling study can be further defined by the quantity or measure of interest, e.g., exceedance frequencies or absolute concentration.

The problem is to relate the chosen performance attributes to specific, available statistics to describe model performance. In doing so, it will be found that certain statistics pertain to several performance attributes. At the same time, each performance attribute generally encompasses several statistical measures. Consequently, using each single statistic as a separate performance attribute is not an acceptable solution because the statistical measures tend to be

> Too redundant--several statistics measure similar or related quantities.

- Too specific--a statistical measure is not general enough to address all of the facets of a performance attribute.

- Too numerous--scores of statistical measures have been designed to describe three or four performance attributes.

Statistical measures, however, do lend themselves to grouping. Upon examination, the AMS workshop report (Fox, 1981) reveals a natural hierarchy of model performance measures and accompanying statistical measures. A hierarchical tree consisting of AMS statistical measures as 'leaves' and performance attributes as major 'branches' appears in figure 2-1. About 25 distinct types of quantitative performance measures are identified for which one or more statistical measures can be used to make a specific statement about model performance and accuracy. In addition, at least eight types of graphical display modes are commonly associated with the AMS set of performance measures. Thus, as can be seen from figure 2-1, summarization is needed to present a clear and unified picture of model performance and accuracy.

The problem then is to group statistics in such a way that they can be interpreted as an aggregate statement about model performance and accuracy. These statements must relate to the model performance attributes and must contain information on

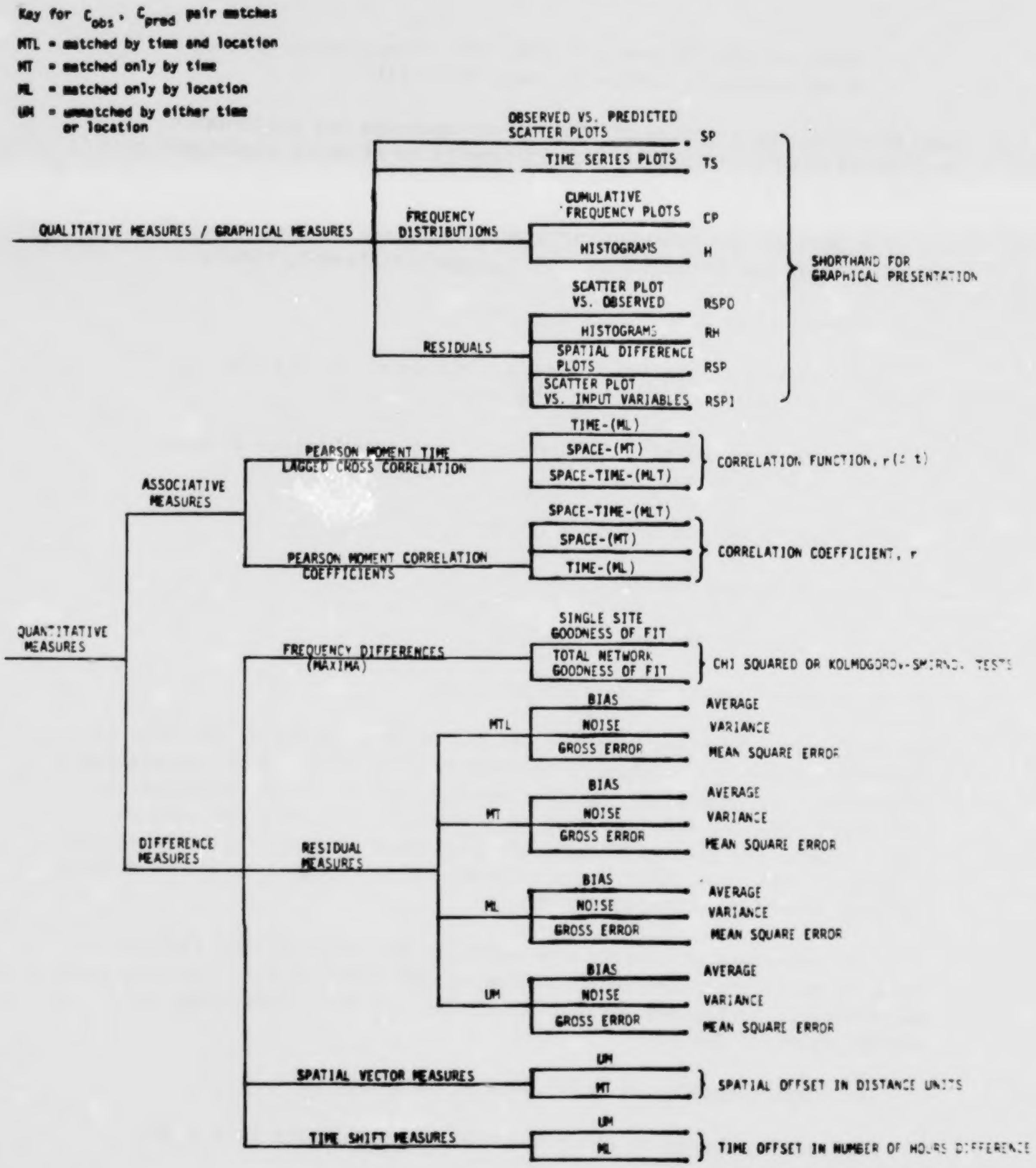
- Absolute confidence--if the evaluation study were repeated 100 times with statistically similar data, into what ranges would the evaluation statistics of 95 of the studies fall?

- Absolute difference (or closeness)--given similar modeling conditions, what differences between predictions and observations would be expected for future model applications?

Several problems are encountered in any attempt to make aggregate statements about model performance and accuracy, among which are the following:

- Specific statements, by themselves quite clear, can conflict with other statements with which they are grouped. The resulting aggregate statement then becomes inconclusive and must be labeled as such.

- If interpretations of specific statistical measures are grouped together, the resulting aggregate statement


Key for C_{obs}, C_{pred} pair matches
MTL = matched by time and location
MT = matched only by time
ML = matched only by location
UM = unmatched by either time or location
QUALITATIVE MEASURES / GRAPHICAL MEASURES
OBSERVED VS. PREDICTED SCATTER PLOTS
SP
TIME SERIES PLOTS
TS
FREQUENCY DISTRIBUTIONS
CUMULATIVE FREQUENCY PLOTS
CP
HISTOGRAMS
H
RESIDUALS
SCATTER PLOT VS. OBSERVED
RSPO
HISTOGRAMS
RH
SPATIAL DIFFERENCE PLOTS
RSP
SCATTER PLOT VS. INPUT VARIABLES
RSPI
SHORTHAND FOR GRAPHICAL PRESENTATION
QUANTITATIVE MEASURES
ASSOCIATIVE MEASURES
PEARSON MOMENT TIME LAGGED CROSS CORRELATION
TIME-(ML)
SPACE-(MT)
SPACE-TIME-(MLT)
CORRELATION FUNCTION, r(τ t)
PEARSON MOMENT CORRELATION COEFFICIENTS
SPACE-TIME-(MLT)
SPACE-(MT)
TIME-(ML)
CORRELATION COEFFICIENT, r
DIFFERENCE MEASURES
FREQUENCY DIFFERENCES (MAXIMA)
SINGLE SITE GOODNESS OF FIT
TOTAL NETWORK GOODNESS OF FIT
CHI SQUARED OR KOLMOGOROV-SMIRNOV TESTS
RESIDUAL MEASURES
MTL
BIAS
AVERAGE
NOISE
VARIANCE
GROSS ERROR
MEAN SQUARE ERROR
MT
BIAS
AVERAGE
NOISE
VARIANCE
GROSS ERROR
MEAN SQUARE ERROR
ML
BIAS
AVERAGE
NOISE
VARIANCE
GROSS ERROR
MEAN SQUARE ERROR
UM
BIAS
AVERAGE
NOISE
VARIANCE
GROSS ERROR
MEAN SQUARE ERROR
SPATIAL VECTOR MEASURES
UM
MT
SPATIAL OFFSET IN DISTANCE UNITS
TIME SHIFT MEASURES
UM
ML
TIME OFFSET IN NUMBER OF HOURS DIFFERENCE

Reproduced from best available copy.

FIGURE 2-1. A HIERARCHICAL TREE OF STATISTICAL PERFORMANCE MEASURES FOUND IN THE AMS WORKSHOP STUDY

Source: Fox (1981)

contains diverse elements that make direct comparison with other aggregate statements very difficult.

Consequently, the grouping of statistical measures and performance evaluation studies into aggregate statements is of major importance to interpretive usefulness.

For each of the models of interest, there are several types of groupings that can be made for each aggregate statement, namely

- > Grouping by statistical measures:
  - Sample stratification (e.g., all cases or only the highest 10 percent of the cases).
  - Various matches of predictions and observations by time and location.
- > Grouping by evaluation studies:
  - Averaging periods
  - Chemical species
  - Model application configurations.

The statistical families representing the major branches of the tree in figure 2-1 appear to be natural boundaries along which to define aggregate statements. In this study, active chemical models are grouped separately from inert chemical models. Major application configurations such as rural, urban, and complex terrain are also separated into major categories of usage. Thus, the differences between evaluation studies within a major category should be acceptably small.

The aggregate statements are meant to represent average tendencies, since it is entirely possible for some of the statistics to conflict with one another. Variations in statistics from one evaluation study to another might be due to

- > Varying quality of input data for the model.
- > Varying degree to which the model is applicable in a given modeling study.
- > Varying use of model options.

> Varying levels of restriction in time-location matches of prediction and observation.

> Varying ways of stratifying the model results.

If a systematic variation such as a variation in the degree of time-location match occurs, this variation should be noted, and an aggregate statement representative of an average or midrange value of the statistical measure should be made. If there is no apparent systematic reason for such a range in a statistical measure, the results will have to be labeled inconclusive. Similarly, the aggregate statements should be considered inconclusive if the confidence interval cannot be calculated. For example, an inability to calculate a confidence interval for bias destroys most of the meaning of the statistics.

Aggregate statements were constructed for eight groups of statistics found in the AMS Woods Hole Workshop (Fox, 1981), namely

> Pearson moment correlation coefficient.

> Time lagged cross-correlation function.

> Goodness of fit tests for cumulative frequency distributions.

> Bias as measured by the average of the residuals.

> Noise as measured by the variance of the residuals.

> Gross error as measured by the root mean square error.

> Vector differences in location between highest observed and predicted concentrations.

> Time difference between the occurrence of the highest observed and predicted concentrations.

The aggregate statement content is to some degree limited by what constitutes substantial or insubstantial quantities. For example, what is an acceptable correlation coefficient--0.6, 0.8, 0.9? Statistical measures of performance attributes such as bias, noise, and gross error are unbounded; thus it would be even more difficult to define substantial levels unless these measures were compared to other statistics such as the moments of the predictions or observations. Recognizing these problems, the AMS workshop participants (Fox, 1981) stated that "It is not appropriate to establish specific numerical standards for model performance...."

The aggregate statements are formulated according to the following basic guidelines:

- > Confidence intervals and statements of significance (e.g., difference from zero or some other meaningful value) at the 95-percentile level are stressed, as they are in the AMS workshop report.
- > Statement of average statistics with appropriate range or confidence intervals.
- > Avoidance of qualitative statements of skill, e.g., weak, strong, slight, good, bad.
- > Avoidance of relative or interpretive measures that do not directly arise from the AMS workshop study.

If such guidelines can be followed, the aggregate statements are apt to be less subject to spurious interpretation. The aggregate statements constructed from the previous considerations are shown in table 2-1.

2.2 RECASTING MODEL EVALUATION STATISTICS INTO THE AMS RECOMMENDED STATISTICS

The major AMS recommended statistics presented in this study are mathematically defined in a glossary appearing at the beginning of this report. Certain pairings of observations with predictions in time or space can generate impossible or meaningless statistics, e.g., correlation coefficients on unmatched but ranked predictions and observations. Therefore, if such meaningless statistics occur in the literature, they are ignored in this study. The graphical forms of presentation suggested by the AMS workshop as well as other useful types of presentations are explained and given a shorthand designation in the glossary. The glossary should provide the reader with sufficient information to interpret the tables compiled for each of the models in succeeding sections.

A number of statistics not recommended by the AMS workshop appear in the model evaluation studies which are referenced in later sections. An attempt will be made to extract AMS recommended statistics from these other measures wherever possible. Although the AMS recommended statistics are of major interest in this study, the non-AMS statistics will occasionally be used as supporting information. Only general comments can be made about many of the non-AMS statistics because their informational content is not easily interpreted.

An effort has been made to lump non AMS statistics into "families" for easier interpretation or recasting into AMS recommended statistics. Some of the most commonly occurring "families" are the following:

- Ratio of predictions and observations. (The ratios are substituted for the original observations and predictions and residuals. The moments, correlation coefficients, and frequency distributions are calculated using these modified variables.) Examples are
    - C_{obs}/C_{pred}
    - C_{pred}/C_{obs}
    - $(C_{obs}-C_{pred})/C_{obs}$
    - $(C_{obs}-C_{pred})/C_{pred}$
    - $2\ (C_{obs}-C_{pred})/(C_{obs} + C_{pred})$.
- The percentage of predictions lying within N percent of the observations.(Similar types of statistics are skill indexes dependent on the number of predictions within a threshold of the observations.) Typical percentages are
    - Percent of total predictions for which $0.5\ C_{obs} < C_{pred} < 2\ C_{obs}$.
- Least square regression statistics where C_{obs} is written as a function of C_{pred}, e.g., the linear case where $C_{obs} = a + b \cdot C_{pred}$ and a, b, and the squared correlation, r^2, are given. Sometimes the standard error of a and b appear also.
- Largest positive and negative errors and range of errors.
- Coefficients of variation of the residuals where the standard deviation of the residuals is transformed by dividing through by the mean observation or prediction.

Several of the families of statistics previously described present unique problems of interpretation. The first family contains problems arising from the use of ratios, namely

- Ratios like C_{obs}/C_{pred} are not very well behaved when the denominator becomes small.

- Statistical measures generated with ratios are difficult to interpret, partly because absolute measures about concentration cannot be obtained.

- Although some ratios may be well behaved, or bounded, the results of an evaluation study cannot be easily compared with those of another unless the actual mean concentrations are known.

The second family, the percentage of predictions lying within a prescribed percent of the observed values, can be misleading if not accompanied by a scatter plot. Such a scatter plot frequently shows a distribution of errors which is non-normal and often multi-modal. For example, if 60 percent of the errors are within prescribed bounds, the remaining errors may cluster near the bounds, or may cluster distant from the prescribed bounds. In the first instance, the model would be predicting better than the measure indicates, while in the second instance the model would be predicting much worse. Other measures, e.g., the standard deviation of the errors would at least partially account for the differences between the two examples.

The family of least squares regression statistics is useful only if information about the goodness of fit is available. One should be aware that the slope and intercept are derived using the least mean square error assumption. The selection of slope and intercept can be very sensitive to the mean square error (MSE), and small differences in the MSE can result in strikingly different slopes and intercepts.

Finally, the largest negative and positive errors are single member sets subject to all the statistical problems associated with outliers. The coefficient of variation, a relatively worthwhile statistic, seems to have been overlooked by the AMS workshop even though it is mentioned by S. Hunter (see Hilst, 1978) as a possible performance measure.

The attempts to recast the non-AMS statistics into AMS statistics met with limited success. Without C_{obs} or C_{pred} or both, the transformation of the ratio measures cannot be accomplished. The measures regarding the percentage of predictions that fall within a set percentage of observations could not be related to any of the AMS recommended measures, and the skill indexes were similarly useless. Also, the largest positive and negative errors did not yield any information on AMS recommended statistics.

Some additional information was provided, however, by the least squares statistics, since for a linear regression fit one has

$$C_{obs} = \overbrace{\bar{C}_o - \bar{C}_p \, r_{op} \left(\frac{\sigma_o}{\sigma_p}\right)}^{A} + \overbrace{r_{op} \left(\frac{\sigma_o}{\sigma_p}\right)}^{B} C_p \quad .$$

Thus, if one has A and B and some simple variables such as σ_o, σ_p, then measures like r_{op}, the correlation coefficient between the predicted and observed, can be retrieved. In the few such cases that occurred, these statistics were retrieved and reported.

Information was also provided by the non-AMS coefficient of variation, which can be used to obtain the mean observation from the standard deviation, and vice versa, when only one of these measures is present in the report.

2.3 SPECIFIC METHODS USED TO CALCULATE STATISTICS

During the course of reviewing the model evaluation studies, several evaluation studies were found to contain few or none of the AMS recommended statistics, although the n highest prediction-observation data pairs were present. In such cases the observation-prediction pairs were put into a statistical program which generates the full set of AMS recommended statistics. Such generated statistics are referenced as coming from the present study. Since these statistical summaries focus only on a selected portion of the total population, the usual assumptions regarding randomness and sample independence are not strictly valid. However, since these statistics are intended primarily as descriptive measures of model peformance, such assumptions are of less importance than if formal hypothesis testing were being used to characterize differences between the observed and predicted concentrations.

Several of the SAI models were evaluated using the original data used in several previous model exercises which are referenced in later sections. In addition the Fisher Z statistic was used to generate 95 percentile bounds for all the correlations reported with an available sample size estimate. If multiple sample sets were reported in a study, the statistics of the sample set with the highest (peak) observations were reported provided the sample size was greater than 20.

Several graphical measures are of use in obtaining AMS recommended statistics. The cumulative frequency distributions were used to estimate the maximum absolute difference between the observed and predicted frequency distributions. With this number the two sample Kolmogorov-Smirnov test, which is defined in many statistical texts such as Lindgren (1976), was calculated if the sample size was available. For small sample sets, scatter plots of observed or predicted concentrations were digitized so that estimates of AMS statistical parameters were generated.

TABLE 2-1. SOME AGGREGATE STATEMENTS OF MODEL PERFORMANCE AND ACCURACY

Performance Attribute	General Statistical Measure	Aggregate Statement
Association of predictions with observations	Pearson moment correlation coefficient with Fisher Z statistic confidence interval	In general, there (is/is not) a significant positive relationship between the predictions and observations at the 95% level; the average and 95% confidence interval of the coefficient is ____ < ____ < ____
Degree of temporal alignment of predictions with observation	Time-lagged cross correlation function with Fisher Z statistic confidence interval	The cross correlation function indicates an average time shift of ___ hr (forwards/backwards) which (is/is not) significant at the 95% level
Degree to which predicted exceedance rates agree with observed exceedance rates	The Kolmogrov-Smirnov D statistic test or the Chi-square tests for goodness-of-fit.	In general, the observed and predicted cumulative frequency distributions (are/are not) identical at the 95% level
The average differences between predictions and observations (bias)	The average residual with students T statistic confidence interval	In general, the model tends to predict (with/without) a bias at the 95% level; the average bias and its 95% confidence interval is ____ ± ____ which indicates (over/under) prediction
The average variation in the difference between predictions and observations (noise)	The variance of the residuals with 95% confidence interval estimated by the chi-square statistic	The average variance and its 95% confidence interval is ____ < ____ < ____ (specie-averaging time), indicating a general (lack of/presence of) model prediction ability for explaining the variability of the observations
Average total variability between the predictions and observations (gross error)	The root mean square error (RMSE) of the residuals on average absolute error (AAE)	The average RMSE and its range is ____ < ____ < ____; the RMSE (is/is not) much different if only the average prediction is used as an estimator

TABLE 2-1 (Concluded)

Performance Attribute	General Statistical Measure	Aggregate Statement
Spatial offset of the locations where similarly ranked predictions and observations occur	Vector difference in location of nth largest observation and its matching prediction (case-by-case)	The average difference in distance and its range is ____ < ____ < ____; the average difference in direction and its range is ____ < ____ < ____; the difference in distance and direction (is/is not) substantial, compared with the space scale of the modeling region
Time offset between occurrences of similarly ranked observations and predictions	Time differences between the occurrence of the nth largest observation and its matching prediction (case-by-case)	The average time offset is a (lag/advance), with an average and range of ____ < ____ < ____

3 RURAL POINT SOURCE MODELS

A group of models has been constructed or adapted for predicting concentrations from point sources in a rural environment. Eleven of these models have been identified for review:

> TCM2

> MPTER

> SCSTER

> TEM8

> ERTAQ

> CRSTER

> PLUME5

> 3141/4141

> MPSDM

> COMPTER

> MULTIMAX.

Information about these models such as developer and outstanding features is given in table 3-1. Note that TCM2 differs most from the other 10 models since it is limited to annual averages.

Several of the models are closely patterned after CRSTER, with the major difference being that an algorithm to treat complex terrain greater than the stack height has been added. Of these models, some, such as PLUME5, 3141/4141, and COMPTER, were evaluated for complex terrain only. To maintain order in the model grouping these evaluations are presented in the section on complex terrain models.

To give the reader an indication of the statistics appearing in the remaining evaluation studies, a summary is provided in table 3-2 for studies in rural, flat, or gently rolling terrain. Of the 11 models, only 3 have been evaluated in flat or rolling terrain. Of these 3, the data from the sole MPSDM study could provide no AMS recommended statistics. The majority of the evaluation studies are for CRSTER.

Table 3-3 presents a description of the evaluation studies for the models TEM8 and CRSTER. In some cases, sample size is not given directly; sometimes estimates must be made. In general, table 3-3 shows that most of the data bases are characterized by

> Less than a dozen monitors, all within 20 km of the stack

> Gently rolling terrain

> Elevated buoyant emission releases.

Statistical measures and aggregate statements for TEM8 appear in table 3-4. It should be noted that the confidence intervals on the bias cannot be calculated because the observed and predicted standard deviations do not appear in the study. Thus, although a consistent tendency for underprediction is shown, it is not possible to determine whether or not the bias is significantly different from zero. This problem appears in most of the model evaluation studies that follow. In addition, the median rather than the average residual is reported for TEM8. Although it is unclear how close the average and the median residuals actually are, they are assumed equivalent. The results indicate that, although the frequency distributions are similar, the degree of paired prediction-observation correlation (hence, causality) is not significantly positive.

The statistical measures and aggregate statements for CRSTER are displayed in table 3-5, subdivided according to averaging period. In several cases, a range of correlations is presented instead of a single average correlation. This was done when several studies showed a range of average correlations obtained on different subsets of the data, for example when correlations were reportd for each stability class. In some studies, such as Londergan, et al. (1980), up to a dozen tracer or monitoring studies are listed. However, on closer inspection many of these studies (such as the surface, nonbuoyant variety) involved uses of the model deemed inappropriate in the Users' Manual. Such studies are not included in this report. Only two of the most appropriate (i.e., elevated release) tracer experiments were chosen. Since the model results are often reported on data which is not representative of the data set as a whole, the following precautions were taken:

> Statistics at one or two stations out of many were not included in calculating the average when they differed substantially from the range of values for the other stations.

> Statistics reported over several stations were averaged and the average is reported.

> Cumulative probability distribution functions (PDFs) at each station were tested if two-thirds of the station PDFs were present; otherwise, the test was dropped.

The 1-hour average results indicate that, although a significant positive association exists between predictions and observations, the frequency distributions are, in general, different. Failure of the Kolmogorov-Smirnov test seems to occur for lower concentrations where a screening threshold of zero or small concentrations have a strong influence on the similarity of observed and predicted PDFs. For goodness of fit tests to be meaningful, some consistent strategy must be adopted to deal with low concentrations. The bias seems to indicate that overprediction generally occurs when the highest concentrations are compared without regard to time matching, whereas underprediction occurs for the total sample set of observations and predictions matched by time or by time and location. The bias at a particular location can be as great as the standard deviation of the residuals in the case of the highest concentrations occurring at a monitor during a measurement program. The noise and gross error statistics were calculated on sample sets of less than 20; too small a sample to produce reliable results in this case.

The 3-hour average statistics are similar to those of 1-hour averages, with the exception that no particular conclusion can be reached about bias. This result is not surprising, since many of the bias values quoted are still from small sample sets. The 24-hour average statistics also seem to be much like the 3-hour average statistics.

The evaluation study of TEM and CRSTER by Bowne et al. (1981) contains several non-AMS recommended statistics, namely

> The percent of predictions within the range of one half to twice the observations.

> Medians and quantities of the observations, predictions, and residuals.

The comprehensive report on models by Londergan et al. (1980) also presents some non-AMS recommended statistics for CRSTER such as

> Ratios of mean predicted over mean observed and the maximum 10 predictions over the maximum 10 observations.

Ratio measures also occur in Mills and Stern (1975), where the ratios of observed to predicted 1-hour average concentrations are presented. Several other non-AMS statistics appearing in the rural point source model evaluations are

> Relative absolute differences, which appear in the CRSTER evaluation study prepared by Mills et al. (1980) and which consist of the absolute value of the residuals divided by the observed.

> Several types of scoring functions for results falling within certain concentration ranges, which occur in the MPSDM evaluation study by ERT (1978).

TABLE 3-1. SUMMARY OF THE VARIOUS RURAL AIR QUALITY MODELS

Name	Developer and Reference	Description
TCM2	Staff of Texas Air Control Board; Texas Air Control Board Austin, TX (Texas Air Control Board [1979])	Multiple point- and area-source Gaussian climatological dispersion model; long-term concentration averages are calculated from joint meteorological frequency distributions
MPTER	T. E. Pierce and D. B. Turner; Environmental Sciences Research Laboratory Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, NC (Pierce and Turner [1980])	Short-term Gaussian diffusion model for multiple point sources; contains options for gently rolling terrain and buoyancy-induced dispersion
SCSTER	M. H. Malik and B. Baldwin Southern Company Services, Inc. Technical Engineering Systems Atlanta, Georgia (Malik and Baldwin [1980])	Multiple point-source short-term Gaussian diffusion model; handles terrain heights greater than stack height; effects of crosswind shear on plume rise are optionally included
TEM8	Staff of Texas Air Control Board; Texas Air Control Board Austin, TX (Texas Air Control Board [1979])	Multiple point-source Gaussian diffusion model; contains power law corrections to P-G dispersion coefficients for longer averaging times
ERTAQ	D. K. Weisenstein and J. H. Wallace; Environmental Research and Technology Concord, MA (ERT [1980a])	Multiple point; line; and area-source Guassian dispersion model; both short-term (hourly) and climatological (inputs of frequency distribution) modes are available. Horizontal dispersion is treated through sector averaging; deposition of particles is included
CRSTER	Monitoring and Data Analysis Division Office of Air Quality Planning and Standards U.S. Environmental Protection Agency Research Triangle Park, NC (EPA [1977])	Gaussian diffusion model for multiple co-located point sources in flat or gently rolling terrain; the receptor grid is azimuthally fixed
PLUME5	J. Hsiung and H. Chase Pacific Gas & Electric Company San Francisco, CA (PG&E [1981])	Multiple point-source Gaussian diffusion model; buoyancy-directional wind-shear-enhanced dispersion, plume penetration through a stable layer and simplified ozone chemistry are treated
3141/ 4141	Enviroplan West Orange, NJ (Enviroplan [1980])	Gaussian diffusion model for multiple co-located point sources; identical to CRSTER but handles complex terrain, buoyancy-enhanced dispersion, and power law correction to P-G dispersion coefficients for longer averaging times
MPSDM	E. Irvine, B. Nitta, and S. Machiraju Environmental Research and Technology Concord, MA (ERT [1980b])	Multiple point-source Gaussian dispersion model with complex terrain handling capability; generates observed-predicted statistics
COMPTER	Alabama Air Pollution Control Commission Montgomery, Alabama (Alabama Air Pollution Control Comission [1980])	Multiple point-source Gaussian dispersion model with various terrain options
MULTIMAX	J. H. Moser Shell Development Company Westhollow Research Center Houston, TX (Moser [1979])	Multiple point-source Gaussian diffusion model with options for treating rolling terrain and power law correction of concentrations (to account for longer averaging times)

TABLE 3-2. AVAILABLE STATISTICAL MEASURES FOR THE RURAL POINT-SOURCE MODELS

Model	Reference	Correlation Coefficients	Frequency Distribution Differences	Residual Measures Bias	Residual Measures Noise	Residual Measures Gross Error	Graphic Display*
TEM8	Bowne et al. (1981)	x	x	x			SP, CP
CRSTER	Bowne et al. (1981)	x	x	x			SP, CP
	Londergan et al. (1980)	x					
	Mills and Record (1975)		x	x	x	x	CP
	Mills and Stern (1975)		x	x	x	x	CP
	Mills et al. (1980)			x	x	x	SP
	Klug (1974)	x					
MPSDM	ERT (1978)						
TCM2	No appropriate studies submitted for review by developers						
MPTER	No appropriate studies submitted for review by developers						
SCSTER	No appropriate studies submitted for review by developers						
ERTAQ	No appropriate studies submitted for review by developers						
PLUME5	No appropriate studies submitted for review by developers						
3141/4141	No appropriate studies submitted for review by developers						
COMPUTER	No appropriate studies submitted for review by developers						
MULTIMAX	No appropriate studies submitted for review by developers						

---

* See glossary for an explanation of abbreviations.

TABLE 3-3. A DESCRIPTIVE SUMMARY OF EVALUATION STUDIES FOR THE RURAL POINT SOURCE MODELS

Reference	Model	Species	Topography	Land Use	Surface Cover	No. of Receptors	Minimum Source-Receptor Distance (km)	Maximum Source-Receptor Distance (km)	Source Type	Averaging Times	Size; Type
[1] Bowne et al. (1981)	TEM8	SO_2	Flat	Agriculture	Corn	28	2	20	Buoyant—187 m	1, 3hr	500-2000; all
	CRSTER									1, 3hr	500; all (time match only) 2000; all, 2000; all (3 hr
[2] Bowne et al. (1981)	TEM8	SF_6	Flat	Agriculture	Corn	~ 200	0.5	50	Buoyant—187 m	1 hr	70; highest
	CRSTER										81; highest
[3] Bowne et al. (1981)	TEM8	SO_2	Flat	Agriculture	Corn	28	2	20	Buoyant--187 m	3 hr	30; highest
[4] Mills and Record (1975)	CRSTER	SO_2	Gently rolling	Rural	Scrub Pine	4	1.4	4.8	1 stack 91 m	1, 3, 24 hr	~ 8000*: all (1-hr), 8; highest (matched by location only)
[5] Mills and Stern (1975); Stuart Plant	CRSTER	SO_2	Gently rolling	Rural-suburban	Unknown	7	2.4	13.3	4 stacks, each 244 m	1, 3, 24 hr	~ 2000*: all (1-hr), 14; highest (matched by location only
[6] Mills and Stern (1975); Muskingum Plant	CRSTER	SO_2	Gently rolling	Rural-suburban	Unknown	4	4.2	19.6	2 stacks, each 251 m	1, 3, 24 hr	~ 1120*: all (1-hr), 8; highest (matched by location only)
[7] Mills and Stern (1975); Philo Plant	CRSTER	SO_2	Gently rolling	Rural-suburban	Unknown	6	1.3	5.6	3 stacks, each 82 m	1, 3, 24 hr	~ 1900*: all (1-hr), 12; highest (matched by location only)
[8] Mills et al. (1980) Clifty Creek	CRSTER	SO_2	Gently rolling	Rural	Woods-farmland (Z_o ~ 20 cm)	6	3.1	15.0	3 stacks, each 208 m	1, 3, 24 hr	Unknown
[9] Mills et al. (1980) Tanners Creek	CRSTER	SO_2	Gently rolling	Rural	Woods-farmland (Z_o ~ 20 cm)	6	1.4	16.1	4 stacks, 83-168 m	1, 3, 24 hr	Unknown
[10] Mills et al. (1980) Muskingum	CRSTER	SO_2	Gently rolling	Rural	Woods-farmland (Z_o ~ 20 cm)	10	1.0	20.0	2 stacks, 252 m	1, 3, 24 hr	Unknown
[11] Mills et al. (1980) Gavin-Kyger-Sporn	CRSTER	SO_2	Gently rolling	Rural	Woods-farmland (Z_o ~ 20 cm)	11	(widely spaced sources)	(widely spaced sources)	6 stacks, 336-164 m,	1, 3, 24 hr	Unknown
[12] Klug (1976)	CRSTER	SO_2	Gently rolling	Rural	?	6	3.2	27.0	3 stacks, 183-244m	1 hr	8760*: all hrs 52,560*: all samples
[13] Londergan et al. (1980) Hanford	CRSTER	various tracers	Flat	Rural	Sagebrush	~ 132	0.2	12.8	Non-buoyant; 111 m	30 min	1300-5300: all
[14] Londergan et al. (1980) Benecia	CRSTER	SF_6	Flat	Rural	Trees	~ 10	16.0	80.0	Buoyant; at 140 m	1 hr	300; all

Note: "All" means the whole sample set.

---

* Estimated

Reproduced from best available copy.

TABLE 3-4. STATISTICAL MEASURES AND AGGREGATE PERFORMANCE STATEMENTS FOR TEM8

(1-hr and 3-hr averaging)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z Statistic for the 95% confidence interval	500-2000 [1] 70 [2] 30 [3]	$-0.39 < -0.34 < -0.28$ [1]	$-0.01 < 0.05 < 0.11$ [1] $-0.40 < -0.18 < 0.06$ [2] $-0.44 < 0.10 < 0.27^{*}$ [3]			Generally, there is not a significant positive relationship occuring between the predictions and observations at the 95% level
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval					Passes the K-S test [1]	The observed and predicted cumulative frequency distributions are, in general, identical at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t\ \alpha\ (S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		20 ppb ± ? [2]	19^{*} ppb ± ? [1] 28^{*} ppt ± ? [2] 10 ppb ± ? [3]			The average bias is 22 ppb for SO_2 and 28 ppt for SF_6, indicating a general underprediction tendency
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residuals as an estimator of σ_d^2 confidence interval $\nu S_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu S_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level						
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error [$RMSE_d$] and the average absolute error $\lvert d \rvert$						

---

* Median reported.

N.B. Numbers in brackets refer to the citations presented in table 3.3.

TABLE 3-5. STATISTICAL MEASURES AND AGGREGATE PERFORMANCE STATEMENTS FOR CRSTER
($\mu g/m^3$)

(a) 1-hr Average

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	500 [1a] 200 [1b] 81 [2a] ~ 8000 [4a] 8 [4b] ~ 2000 [5a] ~ 1120 [6a] ~ 1900 [7a] 14 [5b]	0.18 < (0.18-0.28) < 0.28 [12b] 0.35 < (0.40-0.63) < 0.66 [13] 0.02 < (0.20-0.38) < 0.51 [14] -0.33 < -0.20 <- 0.06 [1b]	0.47 < (0.47-0.65) < 0.65 [12b] -0.19 < 0.20 < 0.40 [2a] 0.09 < 0.18 < 0.26 [1a]			A significant, positive relationship between the observations and predictions at the 95% level; the average correlation is 0.30
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval	8 [6b] 12 [7b] Unknown [8] Unknown [9] Unknown [10] Unknown [11] ~ 8760 [12a] ~ 52,560 [12b] 1300-5300 [13] 300 [14]				Fails K-S test [4a] Fails K-S test [5a] Fails K-S test [6a] Fails K-S test [7a] Fails K-S test [2a]	The observed and predicted cumulative frequency distributions are, in general, not identical at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha}(S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		4.75 ± ? [4a] 0.84 ± ? [8] 23 ppb ± ? [1b]	61 ppt ± ? [2a] 26 ppb ± ? [1a]	229 ± ? [4b] -112 ± ? [5] -68 ± ? [6] -1375 ± ? [7]	-46 ppb ± ? [1a]	Inconclusive
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance with residual as an estimator of σ_d^2 confidence interval $\nu S_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu S_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level		? < 69^2 < ? [8]		$51^2 < 71^2 < 133^2$ [4b] $348^2 < 444^2 < 710^2$ [5b] $231^2 < 320^2 < 512^2$ [6b] $870^2 < 1128^2 < 1635^2$ [7b]		The average variance and its 95% confidence interval is $483^2 < 627^2 < 818^2$
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error [$RMSE_d$] or the average absolute error $\overline{\lvert d \rvert}$		$\overline{\lvert d \rvert}$ = 23 [8] $RMSE_d$ = 69 [8]		$RMSE_d$ = 240 [4b] $RMSE_d$ = 458 [5b] $RMSE_d$ = 327 [6b] $RMSE_d$ = 1778 [7b]		The average $RMSE_d$ and its range is 240 < 700 < 1778

TABLE 3-5 (Continued)
($\mu g/m^3$)

(b) 3-hour Average

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	122 [1] ~ 2000 [5a] ~ 1120 [6a] ~ 1900 [7a]	$-0.33 < -0.16 < 0.02$ [1]	$-0.04 < 0.22 < 0.38$ [1]			Inconclusive
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval	14 [5b] 8 [6b] 12 [7b] 50 [8] 50 [9] 50 [10] 50 [11]				Fails K-S test [5a] Fails K-S test [6a] Fails K-S test [7a]	The observations and predicted cumulative frequency distributions are, in general not identical at 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} [S_o^2/N + S_p^2/N]^{1/2}$ where α = 95% confidence level			18 ppb ± ? [1]	122 ± ? [5b] 163 ± ? [6b] -808 ± ? [7b]	-9.1 ± ? [8] -568 ± ? [9] 156 ± ? [10] 193 ± ? [11] 13 ppb ± ? [1a]	Inconclusive
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residuals as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$ where α = 95% confidence level				$277^2 < 353^2 < 496^2$ [5b] $150^2 < 208^2 < 333^2$ [6b] $540^2 < 700^2 < 1015^2$ [7b]		The average variance and its 95% confidence interval is $360^2 < 468^2 < 680^2$
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error [$RMSE_d$] and average absolute error $\lvert \bar{d} \rvert$				$RMSE_d = 373$ [5b] $RMSE_d = 264$ [6b] $RMSE_d = 1069$ [7b]		The average $RMSE_d$ and its range is $264 < 568 < 1069$

TABLE 3-5 (Concluded)

($\mu g/m^3$)

(c) 24-hour Average

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location): Time and Location	Time Only	Location Only	Unmatched	Aggregate Statement of Model Performance and Accuracy
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95 % confidence interval	~ 333 [a] ~ 83 [5a] ~ 46 [6a] ~ 79 [7a]					
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval	8 [4b] 14 [8b] 8 [6b] 12 [7b] 50 [8] 50 [9] 50 [10] 50 [11]				Fails K-S test [4a] Fails K-S test [5a] Fails K-S test [6a] Fails K-S test [7a]	The observed and predicted cumulative frequency distributions are, in general, not identical at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} [S_o^2/N + S_p^2/N]^{1/2}$ where α = 95% confidence level				41 ± ? [4b] 50 ± ? [5b] 75 ± ? [6b] -133 ± ? [7b]	34.5 ± ? [8a] -45.6 ± ? [9] 84.8 ± ? [10] 75.2 ± ? [11]	Inconclusive
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as and estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level				$12^2 < 16^2 < 27^2$ [4b] $43^2 < 55^2 < 77^2$ [5b] $25^2 < 35^2 < 56^2$ [6b] $103^2 < 134^2 < 194^2$ [7b]		The average variance and its 95% confidencce interval is $57^2 < 75^2 < 108^2$
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error [$RMSE_d$] or the average absolute error $\lvert\bar{d}\rvert$				$RMSE_d$ = 44 [4b] $RMSE_d$ = 74 [5b] $RMSE_d$ = 82 [6b] $RMSE_d$ = 188 [7b]		The average $RMSE_d$ and its range is 44 < 97 < 188;

Reproduced from best available copy.

4 URBAN AIR QUALITY MODELS

The following six urban air quality models are considered in this study:

> AQDM
> CDM
> ERTAQ
> RAM
> TCM2
> TEM8

Brief descriptions of these models can be found in table 4-1. From this table, one can see that several of the models operate in a long-term climatological mode: instead of hourly or even daily meteorological parameters, this mode uses only the frequency distributions of meteorological parameters as input. Three of the models--AQDM, CDM, and TCM-2--operate in this configuration. A fourth, ERTAQ, can be optionally exercised in such a climatological mode. RAM and TEM8 do not operate in a climatological mode, but use hourly meteorological data.

Evaluation studies for each model are compiled in table 4-2, and the AMS recommended statistical groups that are given in each study are indicated. The most prevalent statistic, reported by all but two of the studies, is the correlation coefficient. Most of the graphical displays were of either the scatter plot or cumulative frequency distribution varieties. Some of the concentrations displayed were reduced by estimated background concentrations while others were not. This and other differences made comparisons difficult.

Table 4-3 summarizes the evaluation studies reviewed for the Urban Air Quality Models. The table lists source and pollutant types, averaging times, terrain (topography, land use, and surface cover), number of samplers and the number of samples. The range of distances between source and receptor is given when available.

The AMS recommended statistics gathered for the AQDM evaluation studies are shown in table 4-4. This table indicates that significantly nonzero positive correlations recur. The gross error statistics are for sulfur dioxide (SO_2) and total suspended particulates (TSP), respectively.

The model CDM was run for averaging times varying from several months to a year. Table 4-3 indicates that the sample size is typically between 30 and 100, which is small but statistically adequate. The statistical measures shown in table 4-5 indicate that significantly nonzero positive correlations occur. In addition, the observed and predicted cumulative distributions seem to be quite similar. The bias seems to be insignificant, much less than the RMSE, and of varying size. On the basis of the limited evaluation statistics available, it seems that AQDM and CDM tend to give similar performance statistics.

The short-term urban air quality model, RAM, was represented by more evaluation studies than any of the other urban models. Table 4-3 shows that, for some of the studies, a rural environment was used. Some caution should be exercised when using the rural study results, since the evaluation is for RAM and not its rural offspring RAMR. Usually, the urban studies contain a dozen or more receptor sites, with the surface cover being mainly concrete or asphalt.

The AMS statistics and aggregate statements for RAM for the short (1-3 hr) averaging periods are presented in table 4-6. Although a significantly nonzero correlation is found between the predictions and observations, the observed and predicted frequency distributions do not appear to be similar. The bias values in table 4-6 are rather large. This may be due, in part, to the fact that these studies were conducted in rural rather than urban areas.

The 24-hour average results in table 4-6b are similar to those for the shorter-term averaging periods except that the RMSE values are smaller and the cumulative distribution of the observations and predictions tend to agree with one another. Several of the cumulative distributions could not be used, since at small concentrations one distribution was influenced by background and the other was not. Thus, the two cumulative curves tend to diverge at small concentrations, and the applicability of the Kolmogorov-Smirnov test is degraded. The difference between the rural and urban biases is illustrated in table 4-6 when comparing reference (12) with references (9), (10), and (11). Both sample sets are for the highest concentrations, and there is a difference factor of 30 or more.

Several non-AMS recommended statistics were encountered in the survey of urban air quality models. The evaluation of CDM and AQDM by Busse and Zimmerman (1973) contains

> Largest positive and negative residuals and residuals for maximum observed and predicted concentrations.

> The slope and intercept of a linear regression fit of predictions and observations.

In addition, the evaluation of CDM by Nieuwstadt and van Dop (1975) contains the ratio of the residual to the observations, which is often called the relative error. The RAM evaluation study by Ruff (1980) presents

> Linear regression coefficients

> Several objective loss functions, such as

  - Percent incorrect predictions with absolute error threshold.

  - Symmetric high-low loss function.

The symmetric high-low loss function is zero-one scoring to determine if the concentration is above some arbitrary concentration. Several other scoring statistics are presented in Ruff (1980); however, they are modifications of the two methods presented here.

TABLE 4-1. SUMMARY OF THE VARIOUS URBAN AIR QUALITY MODELS

Name	Developer and Reference	Description
TCM-2	Staff of Texas Air Control Board; Texas Air Control Board Austin, TX (Texas Air Control Board [1981])	Multiple point- and area-source Gaussian climatological dispersion model; long-term concentration averages are calculated using joint meteorological frequency distributions
AQDM	Staff of TRW Systems Group for Department of Health, Education and Welfare, the Public Health Service, the Consumer-Protection and Environmental Health Service, and the National Air Pollution Control Administration Washington, D.C. (TRW [1969])	Multiple point- and area-source Gaussian plume model for calculating long-term (1-yr) average concentrations in urban areas; frequency distributions of meteorological parameters are used as input; a simple statistical analysis package is included
CDM	A. D. Busse and J. R. Zimmerman; National Environmental Research Center Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, NC (Busse and Zimmerman [1973])	Multiple point- and area-source Gaussian plume dispersion model for calculating long-term (1-yr) averages; data input is similar to that of the AQDM
ERTAQ	D. K. Weisenstein and J. H. Wallace; Environmental Research and Technology Concord, MA (ERT [1980a])	Multiple point-, line-, and area-source Gaussian dispersion model; both short-term (hourly) and long-term (climatological) modes of operation are available; horizontal dispersion is treated through sector averaging; enhanced dispersion for urban applications; dry deposition is treated
RAM	D. B. Turner and J. H. Novak; Environmental Sciences Research Laboratory Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, NC (Turner and Novak [1978])	Multiple point- and area-source Gaussian diffusion model for urban applications and short-term averaging periods
TEM8	Staff of Texas Air Control Board; Texas Air Control Board Austin, TX (Texas Air Control Board [1979])	Multiple point-source Gaussian diffusion model; contains power law corrections to P-G dispersion coefficients for longer averaging times

---

* See the glossary for an explanation of the abbreviations.

TABLE 4-2. AVAILABLE STATISTICAL MEASURES FOR THE URBAN AIR QUALITY MODELS

Model	Reference	Correlation Coefficients	Frequency Distribution Differences	Residual Measures			Graphic Display
				Bias	Noise	Gross Error	
AQDM	Turner, Zimmerman, and Busse (1973)	x				x	
	Mc Nidar (1977)	x					SP
CDM	Koch and Fisher (1973)	x		x		x	SP
	Nieuwstadt and Van Dop (1975)	x	x	x	x		SP, CP
	Calder (1973)	x					SP
	Turner, Zimmerman, and Busse (1973)	x	———			x	
RAM	Goldberg and Kern (1978)						
	Kummler et al. (1979)	x	x				CP
	Ellis, Liu, and Dalzel (1980)	x		x			SP
	Morgenstern, Geraghty, and McKnight (1979)		x				CP
	Ruff (1980)	x	x			x	CP, SP
	Londergan et al. (1980)	x					
	Hodanbasi and Peters (1981)		x	x	x	x	SP, CP

TABLE 4-2 (Concluded)

Model	Reference	Correlation Coefficients	Frequency Distribution Differences	Residual Measures			Graphic Display*
				Bias	Noise	Gross Error	
ERTAQ	No appropriate studies submitted for review by developers						
TCM2	No appropriate studies submitted for review by developers						
TEM8	No appropriate studies submitted for review by developers						

TABLE 4-3. SUMMARY OF EVALUATION STUDIES OF URBAN AIR QUALITY MODELS

Reference	Model	Species	Topography	Land Use	Surface Cover	No. of Receptors	Minimum Source-Receptor Distance (km)	Maximum Source-Receptor Distance (km)	Source Type	Averaging Times	Sample Size-Type
[1] Turner, Zimmerman, and Busse (1973)	AQDM	SO$_2$, TSP	Flat	Urban	Skyscrapers	SO$_2$--75 TSP--113	Unknown	Unknown	Point, area	1 yr	75; all 75; all
[2] McNider (1977)	AQDM	TSP	Flat	Urban	Tall buildings	17	Unknown	Unknown	Point, area	1 yr	14; all
[3] Turner, Zimmerman and Busse	CDM	SO$_2$, TSP	Flat	Urban	Skyscrapers	SO$_2$--75 TSP--113	Unknown	Unknown	Point, area	1 yr	75; all
[4] Nieuwstadt and Van Dop (1975)	CDM	SO$_2$	Flat	Urban	Buildings, Water	55	Unknown	Unknown	Point, area	6 mo	55; all
[5] Calder (1973)	CDM	SO$_2$	Flat	Urban	Buildings	40	Unknown	Unknown	Point, area	3 mo	40; all
[6] Goldberg and Kern (1978), Michigan City	RAM	SO$_2$	Flat	Urban	Unknown	4	0.7	7.8	One major point	24 hr	Unknown
[7] Goldberg and Kern (1978), Indianapolis	RAM	SO$_2$	Flat	Urban	Unknown	1	Unknown	Unknown	97 points	24 hr	Unknown
[8] Kummler et al. (1979)	RAM	SO$_2$	Flat	Urban	Unknown	14	Unknown	Unknown	Point, area	3, 24 hr	784; all (1-hr) 98; all (24-hr)
[9] Ellis, Liu, and Dalzel (1980), Eastlake Plant	RAM	SO$_2$	Flat	Rural	Unknown	4	1.9	6.5	Point	3, 24 hr	64; highest
[10] Ellis, Liu, and Dalzel (1980), Avon Lake	RAM	SO$_2$	Flat	Rural	Unknown	3	3.1	3.7	Point	3, 24 hr	48; highest
[11] Ellis and Liu (1981) Wabash River	RAM	SO$_2$	Flat	Rural		4	1.3	6.2	Point	3, 24 hr	64; highest
[12] Hodanbosi and Peters (1981)	RAM	SO$_2$	Flat	Urban	Unknown	33	Unknown	Unknown	Point, area	24 hr	66; highest
[13] Morgenstern, Geraghty, and McKnight (1979)	RAM	SO$_2$	Flat	Urban	Unknown	10	Unknown	Unknown	51 point	3, 24 hr	Unknown (3; hr) <3650-all (24; hr)
[14] Ruff (1980)	RAM	SO$_2$	Flat	Urban	Unknown	13	Unknown	Unknown	Area, point	1, 3, 24 hr. and 1 year	~ 89,000; a (1-hr) ~ 38,000; a (3-hr) ~ 4750; all (24-hr)
[15] Londergan et al. (1980) St. Louis	RAM	FP	Flat	Urban	Unknown	~ 36	1.9	17	Ground-level, non-buoyant point	1 hr	396; all
[16] Londergan et al. (1980) Paramount	RAM	SF$_6$	Flat	Urban	Unknown	~ 13	0.25	1.0	Ground-level, non-buoyant point	1 hr	91; all
[17] Londergan et al. (1980) Goodyear	RAM	SF$_6$	Flat	Urban	Unknown	~ 18	1.0	3.5	Buoyant point, 76 m stack	1 hr	216; all

TABLE 4-4. STATISTICAL MEASURES AND AGGREGATE PERFORMANCE STATEMENTS FOR AQDM
($\mu g/m^3$--annual average)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence level	75 [1a] 75 [1b] 14 [2]	$0.83 < 0.89 < 0.93$ [1a] $0.46 < 0.62 < 0.74$ [1b] $0.31 < 0.72 < 0.91$ [2]				A significant, positive relationship occurs between the observations and predictions at the 95% level; the average correlation and its 95% confidence level is $0.53 < 0.74 < 0.86$
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval						
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} (S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level						
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residuals as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level						
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ or the average absolute gross error $\overline{\lvert d \rvert}$		$RMSE_d = 121$ [1a] $RMSE_d = 36$ [1b] $\overline{\lvert d \rvert} = 92$ [1a] $\overline{\lvert d \rvert} = 28$ [1b]				None--see the gross error statistics

---

N.B. Numbers in brackets refer to the citations presented in table 4-3.
[1a] for SO_2 results
[1b] for TSP results

TABLE 4-5. STATISTICAL MEASURES AND AGGREGATE PERFORMANCE STATEMENTS FOR CDM ($\mu g/m^3$--annual average)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	75 [3] 55 [4] 40 [5]	$0.76 < 0.84 < 0.90$ [3] $0.65 < 0.78 < 0.87$ [4] $0.60 < 0.77 < 0.87$ [5]				In general, significant positive relationship occurs between the observations and predictions at 95% levels; the average correlation and its 95% confidence level is $0.67 < 0.79 < 0.88$
Frequency distribution differences	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval					Passes the K-S test [4]	In general, observed and predicted cumulative frequency distributions are identical at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} [S_o^2/N + Sp^2/N]^{1/2}$ where α = 95% confidence level		-3 ± 7 [3] 6 ± 6.3 [4]				Tends to predict with a bias at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$ where α = 95% confidence level		$160 < 185 < 220$ [4]				None--see the single statistic
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ or the average absolute error $\overline{\lvert d \rvert}$		$RMSE_d = 52$ [3] $\overline{\lvert d \rvert} = 37$ [3]				None--see the single statistic

---

N.B. Numbers in brackets refer to the citations presented in table 4-5.

TABLE 4-6. STATISTICAL MEASURES AND AGGREGATE PERFORMANCE STATEMENTS FOR RAM
($\mu g/m^3$)

(a) 1- and 3-hour Average

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	784 [8] 64 [9] 48 [10] 64 [11] ~ 89,000 [14a] ~ 38,000 [14b] 396 [15]	0.67 < 0.72 < 0.76 [11] 0.67 < 0.80 < 0.88 [10] 0.59 < (0.73 - 0.93) < 0.96 [15] 0.36 < 0.42 < 0.48 [8] -0.03 < 0.22 < 0.44 [9]				In general, a significant positive relationship occurs between the observations and predictions at 95% level; the average correlation and its 95% confidence level is 0.51 < 0.67 < 0.75
Frequency distribution differences	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval					Fails the K-S test [8]	In general, observed and predicted cumulative frequency distributions are not identical at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha}[S_o^2/N + S_p^2/N]^{1/2}$ where α = 95% confidence level				-1068 ± ? [9] -415 ± ? [10] -701 ± ? [11]		Average bias is -728, indicating a general tendency for overprediction; the nonzero significance is unknown
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$ where α = 95% confidence level						
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ and the average absolute error $\overline{\|d\|}$		$\overline{\|d\|}$ = 39 [14a] $RMSE_d$ = 98 [14a] $RMSE_d$ = 77 (3 hr) [14b]				None--see statistic

TABLE 4-6 (Concluded)
($\mu g/m^3$)
(b) 24-hr Average

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence level	Unknown [6] Unknown [7] 98 [8] 64 [9] 48 [10] 64 [11] 66 [12] 4758 [14]	$0.52 < 0.65 < 0.75$ [8]				In general, a significant positive relationship occurs between the observations and predictions at 95% levels; average correlation and its 95% confidence level is $0.52 < 0.65 < 0.75$
Frequency distribution differences	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic for the 95% confidence interval					Inconclusive [6] Inconclusive [7] Passes the K-S test [14]	In general, observed and predicted cumulative frequency distributions are identical at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} [S_o^2/N + S_p^2/N]^{1/2}$ where α = 95% confidence level				$-385 \pm ?$ [9] $-164 \pm ?$ [10] $-246 \pm ?$ [11] $-4.15 \pm ?$ [12]		Average bias is-200, indicating a general tendency for overprediction; the nonzero significance is unknown
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu S_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu S_d^2/\chi^2_{\nu,1-\alpha}$ where α = 95% confidence level				$74^2 < 84^2 < 98^2$ [12]		None--see the single statistic
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-squre error $RMSE_d$ or the average absolute error $\overline{\lvert d \rvert}$		$RMSE_d = 55$ [14]		$\overline{\lvert d \rvert} = 57$ [12] $RMSE_d = 84$ [12]		The average $RMSE_d$ and its range is $55 < 70 < 84$

N.B. Numbers in brackets refer to citations presented in table 4-3.

5 INDUSTRIAL COMPLEX SOURCE MODELS

Some industrial sites, such as refineries, smelters, or other heavy industry, have very complicated source structures. Multiple stacks may be aligned such that a line source effect occurs. If the sites have large buildings and short stacks, building wake effects can be important. Several models have been developed to specifically model such sources. In the present study five such models were identified:

> ISC(LT)

> ISC(ST)

> PASDIM

> BLP

> PAL

A few details about the models such as developer, development date, and general description can be found in table 5-1. One of the models, ISC(LT) operates only in a long-term (climatological) averaging mode. A second model, PASDIM, can be operated in either a short- or long-term averaging mode. All models, with the exception of PAL, have some sort of building-modified plume rise, or dispersion algorithm.

A summary of the model performance studies and the statistical measures calculated in each study are given in table 5-2. From the table it can be seen that only ISC(LT) has not been evaluated. The availability of the statistics indicates that the residual measures are very well represented, partly because new data were taken from the evaluation studies to calculate the AMS recommended statistics. This procedure was done for three of the models: ISC(ST), PASDIM, and BLP.

A description of two of the model applications for the ISC(ST) model is given in table 5-3. The measurement programs are carried out near the source, as one would expect if the effect of stack downwash were being studied. The terrain is not complex, and except for the industrial plant proper, the land is generally rural. The second reference, Bowers and Anderson (1981), was analyzed by taking the data from a tabulation given in the study.

The statistical results of the evaluation studies are presented in table 5-4. In general, the results indicate a lack of significant nonzero association between the predictions and observations either by rank or by time-location match. One study shows some significant bias towards underprediction; the other study, which located closer to the stack, suggests some overprediction, but contains insufficient information to determine whether this overprediction is significant. The SO_2 experiment data was analyzed in the present report according to the methods outlined in section 2.

Table 5-3 describes the source of the data for the single PASDIM evaluation study. The receptor network is not located near the source, and the averaging time chosen was annual. The results shown for this study were calculated from data on a detailed scatterplot; the error in estimating the data from the scatterplot is ± 1 $\mu g/m^3$. The results given in table 5-5 indicate a lack of bias. The degree of association between predictions and observations is substantial despite the small sample size of only 7 to 11 numbers. The cumulative distributions were determined to be identical, using a modification of the K-S test for small (<20) sample size. The degree of agreement between predicted and observed is believed to be partly due to the long-term averaging.

Several line source tracer experiments have been used to evaluate PAL (see table 5-3). One was an elevated release, as from a refinery, and the second was a near ground-level release. The terrain is flat in both cases, and a flourescent particle tracer was used. In addition, PAL was exercised in a rural environment. The results in table 5-6 seem to show that, although a significant positive association occurs between predictions and observations, the elevated source results indicate a significant degree of correlation.

Two evaluation study descriptions for BLP are presented in table 5-3. Both references involve a near source array that would be expected if the effect of stack downwash were important. The samples of the SO_2 data base are limited spatially to two locations; consequently, the data might not be considered spatially representative enough to be used in an evaluation study. Nevertheless, the published top 20 values are used to compute the AMS recommended statistics.

The AMS recommended statistics for BLP are shown in table 5-7. The correlation coefficients in that table were reported in the reference; however, the sample size could only be inferred from statements made in the evaluation study. The results indicate that the statistical accuracy increased as the averaging period increased from 1 to 24 hours. At 1-hour averaging intervals, the model seems to overpredict at the highest concentrations even when the observed and predicted concentrations are matched only by rank. At 24-hour averaging periods, however, the bias becomes insignificant, and the gross error is about a factor of 5 less than the average observed concentrations.

The ISC(ST) evaluation studies by Cramer et al. (1976) contain some non-AMS recommended statistics. The most prominent statistic is the ratio of predicted to observed concentrations. The evaluation studies for BLP also include ratios of concentration.

TABLE 5-1. A SUMMARY OF BASIC INFORMATION ABOUT THE VARIOUS INDUSTRIAL COMPLEX SOURCE MODELS

Name	Developer and Reference	Description
ISC (LT)	J. F. Bowers, J. R. Bjorklund, and C. S. Cheny H. E. Cramer Company, Inc. University of Utah Research Park Salt Lake City, UT (Bowers, Bjorklund, and Cheney [1979])	A multiple source Gaussian climatological dispersion model for calculating long-term averages from frequency distributions; deposition and building wake effects are included
ISC (ST)	Same as ISC (LT)	Short-term multiple point-source Gaussian model; deposition and building wake effects are included
PASDIM	C. L. Norton and A. Roffman Energy Impact Associates Pittsburgh, PA (Norton and Roffman [1980])	Point-, area-, and line-source Gaussian model for calculating short-(1-24 hr) and long-(annual) term averages; particle setting, deposition; stack and building downwash, and potential flow wind model for complex terrain are included
PAL	W. B. Petersen Environmental Protection Agency Research Triangle Park, North Carolina (Petersen [1978])	Point-, area-, and line-source Gaussian model
BLP	L. L. Schulman and J. S. Scire Environmental Research and Technology, Inc. Concord, Massachussetts (Schulman and Scire [1980])	Multiple point- and line-source Gaussian model that includes building downwash with plume rise modification, pollutant decay, and terrain plume height adjustment

TABLE 5-2. AVAILABLE STATISTICAL MEASURES FOR EACH OF THE INDUSTRIAL COMPLEX SOURCE MODELS

Model	Reference	Correlation Coefficients	Frequency Distribution Differences	Residual Measures			Space-Time Offsets	Graphic Measures
				Bias	Noise	Gross Error		
ISC* (ST)	Abrams (1979); Bowers and Anderson (1981) Scire and Schulman (1981)	x	x	x	x	x		CP, SP
PASDIM*	Roffman (1979)	x	x	x	x	x		CP, SP
PAL	Londergan et al. (1980)	x						
BLP	Schulman and Scire (1980);	x	x	x	x	x		CP, SP
	Scire and Schulman (1981)			x	x	x		
ISC (LT)	No appropriate studies submitted for review by developers							

---

* Statistics were generated from tables of observations and predictions.

TABLE 5-3. A DESCRIPTIVE SUMMARY OF EVALUATION STUDIES FOR THE INDUSTRIAL COMPLEX SOURCE MODELS

Reference	Model	Species	Topography	Land Use	Surface Cover	No. of Receptors	Minimum Source-Receptor Distance (km)	Maximum Source-Receptor Distance (km)	Source Type	Averaging Times	Sample Size-Type
[1] Scire and Schulman (1981)	ISC(ST)	SF_6	Gently rolling	Industrial in rural surroundings	Vegetation	19	0.75	2.25	Buoyant line source	1 hr	532; all
[2] Bowers and Anderson (1981), Abrams (1979)	ISC(ST)	SO_2	Flat	Rural	Vegetation	9	1	4	Multiple buoyant point sources	1 hr	143; all
[3] Roffman (1979)	PASDIN	SO_2, TSP	Rolling terrain	Industrial	Unknown	12	9	~ 30	Multiple buoyant point source	Annual	7; all 11; all
[4] Londergan et al. (1980) Fort Wayne	PAL	FP	Flat	Urban to rural	Trees buildings	227	1.6	14	Infinite line source at 91 and 214 m	1 hr	1115–4336; all
[5] Londergan et al. (1980) Woodlot	PAL	FP	Flat	Rural	Trees and fields	209	0.3	5.5	Line source 2 m	1 hr	3750; all
[6] Londergan et al. (1980) Hanford (elevated source)	PAL	various	Flat	Rural	Sagebrush	132	0.20	12.8	Non-buoyant 111 m point source	30 min	1300 - 5300; all
[7] Londergan et al. (1980) Benecia	PAL	SF_6	Flat	Rural	Trees	10	16.	80.	Buoyant point source	1 hr	300; all
[8] Scire and Schulman (1981)	BLP	SF_6	Gently rolling	Industrial in rural surroundings	Mixed	19	0.75	2.25	Buoyant line source	1 hr	532; all
[9] Schulman and Scire (1980)	BLP	SO_2	Gently rolling	Industrial in rural surroundings	Mixed	2	0.3	0.6	Buoyant line source	1, 3, 24 hr	20; highest

TABLE 5-4. STATISTICAL MEASURES AND AGGREGATE PERFORMANCE STATEMENTS FOR ICS(ST)
($ppb(SO_2)$ or 10^{-7} $s/m^3(SF_6)$--1 hour average)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient r with Z statistic for the 95% confidence interval	532 [1] 143 [2]	$-0.10 < 0.07 < 0.23$ [2]				In general, no significant, positive relationship occurs between the observations and predictions
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval					Fails K-S test [2]	Observed and predicted cumulative frequency distributions are not identical at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha}(S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		$-10.05 \pm ?$ [1] 23.5 ± 38 [2]				Inconclusive
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu S_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu S_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level		$13.7^2 < 14.4^2 < 15.1^2$ [1] $121^2 < 133^2 < 147^2$ [2]				None-see single statistic
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ or the average absolute error $\|\bar{d}\|$		$\|\bar{d}\| = 12.6$ [1] $\|\bar{d}\| = 235$ [2] $RMSE_d = 270$ [2] $RMSE_d = 17.5$ [1]				None--see single statistic

---

N.B. Numbers in brackets refer to the citations presented in table 5-3.

[1] SF_6 statistic.

[2] SO_2 statistic.

TABLE 5-5. STATISTICAL MEASURES AND AGGREGATE PERFORMANCE STATEMENTS FOR PASDIM
($\mu g/m^3$--annual average)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient r with Z statistic for the 95% confidence interval	7 [3a] 11 [3b]	$-0.01 < 0.75 < 0.96$ [3a] $0.72 < 0.92 < 0.98$ [3b]				In general, a significant, positive relationship occurs between the observations and predictions at the 95% level; average correlation and its 95% confidence level is $0.36 < 0.84 < 0.97$
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval					Passes the K-S test [3a] Passes the K-S test [3b]	The observed and predicted cumulative frequency distributions are identical at the time of 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} (S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		-6.0 ± 24.0 [3a] 7.54 ± 17.31 [3b]				The bias is not significant at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residuals as an estimator of σ_d^2 confidence interval $\nu S_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu S_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level		$12^2 < 17^2 < 30^2$ [3a] $9.6^2 < 13^2 < 20^2$ [3b]				None-see statistical measure
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ or the average absolute error $\lvert\bar{d}\rvert$		$RMSE_d = 18.0$ [3a] $\lvert\bar{d}\rvert = 12.0$ [3b] $RMSE_d = 15.0$ [3a] $\lvert\bar{d}\rvert = 10.0$ [3b]				None-see statistical measure

---

N.B. Numbers in brackets refer to the citations presented in table 5-5.

[3a] SO_2 statistic.

[3b] TSP statistic.

TABLE 5-6. STATISTICAL MEASURES AND AGGREGATE PERFORMANCE STATEMENTS FOR PAL

(a) Line Source Studies

($\mu g/m^3$--1 hour average)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient r with Z statistic for the 95% confidence interval	115-4336 [4] 3750 [5]	0.04 < (0.07-0.10) < 0.13 [4] 0.33 < (0.35-0.56) < 0.59 [5]				A significant, positive relationship occurs between the observations and predictions at the 95% level; the average correlation and its 95% confidence level is 0.17 < 0.27 < 0.36
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval						
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} (S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level						
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level						
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ or the average absolute error $\overline{\lvert d \rvert}$						

TABLE 5-6 (Concluded)

(b) Rural Point Source Studies

($\mu g/m^3$—1 hour average)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location) Time and Location	Time Only	Location Only	Unmatched	Aggregate Statement of Model Performance and Accuracy
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	1300-1500 [6] 300 [7]	0.34 < (0.39-0.63) < 0.66 [6] 0.08 < (0.19-0.33) < 0.43 [7]				In general, a significant positive relationship occurs between the observations and predictions at the 95% level. The average correlation is).38
Frequency distribution difference	Goodness-of-fit—chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval						
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha}[S_o^2/N + S_p^2/N]^{1/2}$ where α = 95% confidence level						
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residuals as an estimator of σ_d^2 confidence interval $\nu S_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu S_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level						
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error [$RMSE_d$] or average absolute error $\lvert\bar{d}\rvert$						

N.B. Numbers in brackets refer to citations presented in table 5-3.

TABLE 5-7. STATISTICAL MEASURES AND AGGREGATE PERFORMANCE STATEMENTS FOR BLP

($\mu g/m^3(SO_2)$ and 10^{-7} $s/m^3(SF_6)$--1, 3, 24 averages)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location): Time and Location	Time Only	Location Only	Unmatched	Aggregate Statement of Model Performance and Accuracy
Associative-correlative	Pearson moment correlation coefficient r with Z statistic for the 95% confidence interval	532 [8] 20 [9a] 20 [9b] 20 [9c]	$0.26 < 0.34 < 0.41$ [9a] $0.0 < 0.44 < 0.74$ [9b] $0.21 < 0.60 < 0.82$ [9c]				A significant, positive relationship occurs between the observations and predictions; the average correlation and its 95% confidence level is $0.16 < 0.46 < 0.66$
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval					Fails k-s test [9a] Passes k-s test [9b] Passes k-s test [9c]	As the averaging period increases, the observed and predicted cumulative frequency distributions become identical at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha}(S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		-1.12 ± 7 [8]			-241.7 ± 48 [9a] -85 ± 54 [9b] -16 ± 21 [9c]	At the shortest averaging periods, a significant negative bias occurs at the 95% level which indicates over predictions
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/{}^2_{\nu,1-\alpha}$ where α = 95% confidence level		$9.0^2 < 9.4^2 < 9.9^2$ [8]			$81^2 < 109^2 < 137^2$ [9a] $51^2 < 66^2 < 87^2$ [9b] $11^2 < 15^2 < 20^2$ [9c]	The variance tends to decrease increasing averaging period
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ or the average absolute error $\overline{\lvert d \rvert}$		$\overline{\lvert d \rvert}$ 5.43 [8] $RMSE_d$ 9.4 [8]			$RMSE_d = 262$ [9a] $\overline{\lvert d \rvert} = 251$ [9a] $RMSE_d = 107$ [9b] $\overline{\lvert d \rvert} = 95$ [9b] $RMSE_d = 17$ [9c]	The measures of gross error tend to decrease with increasing averaging period

---

N.B. Numbers in brackets correspond to the citations presented in table 5-9.

[9a] 1 hr SO_2
[9b] 3 hr SO_2
[9c] 24 hr SO_2

Reproduced from
best available copy.

6 REACTIVE CHEMISTRY MODELS

Because many pollutants released into the atmosphere are not inert, their chemical transformations often result in secondary pollutants having greater potential for harmful environmental or health effects. Several models designed to simulate these chemical transformations are considered in this section. These models are

> ELSTAR (Environmental Lagrangian Simulator of Transformation Atmospheric Reaction).

> SAI RPM-II (Reactive Plume Model).

> SAI Airshed.

The model developers and an overview of the model attributes are presented in table 6-1. From this table it can seen that, though these models share a detailed chemical kinetic treatment, they are not applicable to the same situations. ELSTAR is designed to simulate reactions within a parcel of air along its trajectory, typically a parcel originating over or upwind of an urban region. The RPM also simulates chemistry within an air parcel; however, the parcel is taken to be a segment of a major point source plume, and the scale of applicability of the RPM is considerably smaller than that of ELSTAR. The Airshed Model is a completely different type of model, whose domain of coverage is spatially fixed, and therefore, it is useful for examining the detailed interaction among many types of emission sources within an urban region.

A summary of the model performance studies and the statistics generated in each study is presented in table 6-2. Within a number of performance evaluation studies for the Airshed Model, several scenarios were simulated. Within each scenario, statistics were often available for samplers stratified by ozone or NO_2 concentration level. The availability of statistics stratified in this manner are not shown in table 6-2; they can be found in the original reports.

The data bases used to evaluate the ELSTAR model are summarized in table 6-3. The initial model development and evaluation efforts were documented in Lloyd (1979). The Los Angeles Reactive Pollutant Program

(LARPP) was specifically established for the evaluation of a trajectory-type reactive model and was utilized in conjunction with the development of ELSTAR.

ELSTAR requires initial concentrations of pollutants as input to the model. These concentrations are, in most cases, not measured. For such cases, the validation procedure recommended by the model developer is to carry out at least two simulations. This permits changes to be made in the choices of initial concentrations of unmeasured pollutants, to avoid errors due to the input of incorrect initial conditions.

The evaluation studies of ELSTAR can be seen from table 6-3 to be limited both in space and time. The tests are limited to a few days, and to a few trajectories scattered over the Los Angeles basin (100 km by 100 km). Although comprehensive meteorological data were collected to run the models, the sample set with which to evaluate the model is rather small.

In both studies of ELSTAR, only statistics on observations and predictions matched both by space and time are presented. Since several species of pollutants such as NO, NO_2, O_3, and CO_2 were involved in the evaluation studies, the AMS-recommended statistics for all of the species were lumped into a comprehensive table (see table 6-4). The evaluation study of Lloyd et al. (1979) dealt mainly with concentration ratios (e.g., C_{pred}/C_{obs}). Few raw data were presented; consequently, some of the statistics and their confidence intervals could not be calculated.

Several aggregate statements can be made from a consideration of table 6-4:

> For all chemical species there is a positive correlation between predictions and observations that is significant at the 95 percent confidence level. The average correlation and its 95 percent confidence interval is $.57 < .62 < .66$.

> As often happens, the lack of significance of the bias may be due to the large scatter evident in the data. However, the sign of the bias estimate is positive in all cases, indicating that if a significant bias should occur, it would tend toward underprediction.

> The root mean square is much larger than the average residual value for the NO and the O_3 results.

An evaluation study of the RPM was documented in Stewart and Liu (1981). The evaluation centered around two scenario simulations, whose specific details concerning the data base are presented in table 6-3. The data used consisted of a dozen or so aircraft traversals of each power plant plume to sample air parcels originating at the plant being studied. The two data bases differed in that

> The Widow Creek study reported plume average O_3 concentrations only.

> The Oak Creek study reported continuously measured concentrations across the plume.

Although no statistics were presented in Stewart and Liu (1981), the evaluation results were presented as either downwind or cross-plume plots of concentration as a function of distance. A grid was superimposed on the plot to extract data for the generation of statistics. Where cross-plume concentrations were derived, the data extraction necessitated "eyeball" averaging over short distance intervals. Although the error associated with this procedure have not been calculated, they are likely to be relatively small and random in nature. Consequently, they should not introduce bias into the statistics. The statistics derived from the Stewart and Liu (1981) publication are presented in table 6-5. The wide ranges of the confidence interval are due to the very limited sample size, so the results should be interpreted with care.

The statistics in table 6-5 can be summarized in the following aggregate statements

> For all chemical species there is a positive correlation between predictions and observations that is significant at the 95 percent confidence level. The average correlation and its 95 percent confidence interval is $0.66 < 0.88 < 0.96$.

> The bias is not significant at the 95 percent level for any of the chemical species.

> The noise is generally much greater than the bias.

The Airshed Model has been evaluated many times using several different data bases. Often, as model improvements are made, new evaluations are conducted using an older data base. Selection of the Airshed Model evaluation studies to be included in this report was governed by two concerns. The most up-to-date evaluation exercises with a wide variety of different application scenarios were selected and are listed in table 6-3. In addition, information on the specific urban region modeled and the available data are also presented in table 6-3.

Because the urban AIRSHED model is data intensive, it requires extensive meteorological and air quality data for a valid simulation. Generally, the model has been applied to urban regions where the occurrence of smog is not uncommon. Usually, these cities have well established air monitoring networks that supply the necessary air quality data for model evaluation. In one study (i.e., Roth, et al. 1981) an air quality network was established specifically to exercise the Airshed Model. This network was not as extensive as those of large cities. Moreover, it had been operating for only a short period when data for the study evaluation were collected. Thus, although relatively large data samples were used to generate the evaluation statistics, the data bases may not be of uniform quality.

As for the other models, the pairing of predictions and observations is always by time and location, and the major remaining stratification is by chemical specie. Table 6-6 presents the evaluation statistics for NO, NO_2, O_3, and CO respectively. Statistics for the Airshed Model are often presented from stratified data samples. In many evaluation scenarios, statistics may be generated for only the high concentration of a particular species. Unlike the AMS recommendation of selecting the highest M measurements, or even a percentage of the highest measurements, the Airshed statistics have been stratified by measurements exceeding some cutoff value. The concentration cutoffs chosen for this study were the lowest possible.

The standard statistical package used for Airshed simulation generated a correlation coefficient, residual bias and noise, measures of gross error, and space-time offsets. However, confidence intervals were not presented in any of the studies. Graphical display of predicted-measured pairs are often in the form of

> Time series at individual stations

> Mean deviation as a function of concentration

> Normalized absolute error as a function of concentration

> Distribution of residual error

> Spatial plots of concentration difference.

The statistical measures in table 6-6 are summarized in the following aggregate statements:

- For NO and O_3 there is a significant positive correlation at the 95 percent level. The averages and the 95 percent confidence levels are

$$NO\text{--}0.24 < 0.50 < 0.74$$
$$O_3\text{--}0.65 < 0.76 < 0.82.$$

- For NO_2 and CO the correlation coefficient results are inconclusive.

- Though the significance is unknown, the NO_2 bias is consistently positive, indicating a possible tendency for model underprediction.

- The average residual fluctuates in sign for the chemical species NO, O_3, and CO. This tends to render any interpretation for these chemicals as inconclusive.

- The average noise and its 95 percent confidence interval are

$$NO\text{--}1.2 < 3.1 < 6.4\ (\text{pphm})^2$$
$$NO_2\text{--}2.4 < 4.2 < 5.3\ (\text{pphm})^2$$
$$O_3\text{--}8.6 < 10.4 < 12.3\ (\text{pphm})^2$$
$$CO\text{--}0.64 < 1.22 < 1.99\ (\text{ppm})^2.$$

- The average root mean square error is

NO--13.4 pphm
NO_2--8.6 pphm
O_3--2.6 pphm
CO--4.9 ppm.

In addition, spatial and temporal differences in the time and location of the unmatched, highest observed and predicted concentrations appear as the last two rows in table 6-6. The results are inconclusive for the spatial vector differences. However, the temporal differences indicate that there is a time lag on the order of 8 hr for the NO_2 predicted maximum to occur. The other chemical species give varying and, consequently, inconclusive results.

TABLE 6-1. SUMMARY OF BASIC INFORMATION ABOUT THE VARIOUS REACTIVE CHEMISTRY MODELS

Name	Developer and Reference	Description
ELSTAR	Environmental Research and Technology, Inc. Concord, Massachusetts (Lloyd et al., 1979)	• Lagrangian model formulation • Uniform size vertical slice of atmospheric divided vertically into an array of cells • Horizontal uniform pollutant concentration in each cell • Emission input into lower cell, limited emission allowed from outside (laterally) modeling domain • Dry deposition parameterization • Chemical mechanism-46 specie, 60-reaction • Gear-type (1971) integration routine (EPISODE)
RPM	Systems Applications, Inc. 101 Lucas Valley Road San Rafael, Calfornia (Stewart and Liu, 1981)	• Lagrangian model formulation • Horizontal array of expanding cells • Gaussian distribution of inert species (cell membranes impermeable) • Reactive species allowed to diffuse between cells at a rate dependent on cell expansion • Within each cell, concentrations are uniform • Chemical Mechanism-Carbon bond Mechanism, 37 species, 65 reactions • Gear-type (1971) integration
AIR-SHED	Systems Applications, Inc. 101 Lucas Valley Road San Rafael, California (Reynolds et al., 1979)	• 3-dimensional grid model with multiple vertical cells • Point source and area source emissions • Ground level and elevated emissions • Pollutant dry deposition parameterization depended on surface characteristics and atmospheric resistance • SHASTA numerical advection scheme • Plume rise algorithm • Mesoscale emission reaction algorithm • Eddy diffusivity is stability, surface roughness, and altitude dependent

TABLE 6-2. SUMMARY OF THE AVAILABILITY OF MODEL EVALUATION STUDIES AND THEIR AMS RECOMMENDED STATISTICS FOR THE REACTIVE MODELS

Model	Reference	Correlation Coefficients	Frequency Distribution Differences	Residual Measures			Space Time Offsets	Graphical Measure
				Bias	Noise	Gross Error		
ELSTAR	Lloyd et al. (1979)	x		x				TS
	Godden et al. (1980)	x		x	x	x		TS
RPM-II	Stewart and Liu (1981)	x		x	x	x		ISP
AIR-SHED	SAI (1981)	x		x		x		
	Roth et al. (1981)	x		x	x	x	x	TS
	DOT (1979)	x		x	x	x	x	TS
	Reid et al. (1980)	x		x	x	x	x	
	Tesche and Burton (1978)			x	x	x	x	
	Tesche et al. (1981)	x		x	x	x	x	

TABLE 6-3. A DESCRIPTIVE SUMMARY OF DATA BASES USED IN THE REACTIVE MODEL EVALUATION

Reference	Model	Location	Date(s)	Program Participants*	Data Base/Measurements Made	Sample Size for Model Evaluation
[1] Lloyd et al. (1979)	ELSTAR	Los Angeles basin One Trajectory simulation each date	Oct. 1, 5, 10, 11, 24, 25, Nov. 7, 1973	Metronic, EPA, NCAA CRC, CARB	Los Angeles Reactive Pollutant Program Aircraft and mobile van measurements of O_3, NO, NO_x, methane, nonmethane, hydrocarons, CO. Tetroon launcher and tracking Lidar measurements UV radiation measurements Meteorological measurements (standard)	132 data points, 7 separate trajectory simulations with varying numbers of data points
[2] Godden et al. (1980)	ELSTAR	Los Angeles basin	Oct. 16, 26, Nov. 2, 1973, June 28, July 24, Nov. 14, 1974	SCAB LARPP participants	South Coast Air Basin routine network measurements of NO, NO_2, O_3, CO Meteorological Measurements (wind speed, direction, temperature) Also LARPP data base utilized (see Lloyd et al. [1979] description above)	11 data points
[3a] Stewart and Liu (1981)	RPM	Widows Creek Power Plant, Northeastern Alabama	23 August 1978	TVA	Widows Creek plume characterization study Total of 6 plume traverses used Horizontal integrated concentrations of NO_x and O_3	6 data points
[3b] Stewart and Liu (1981)	RPM	Oak Creek Power Plant, Eastern Wisconsin	20 July 1977	MAP3S	Airborne sampling at three downwind locations (32, 82, 124 hr) Continuous sampling of NO, NO_2, O_3	18 data points
[4] Systems Applications, Inc., (1981)	AIRSHED	St. Louis, MI	13 July, 1976	RI	Regional Air Pollution Study (RAPS) data base with RAMS network utilized (Regional Air Monitoring System) 25 sites, 21 within AIRSHED modeling region NO, NO_2, O_3, CO total hydrocarbon, grab bag sample, particulates and meteorological variable were measured at all sites	NO_2 - 294 O_3 ~ 280
[5] Roth, et al. (1981)	AIRSHED	Stockton, CA	13 Sept., 1979	PGE, SRI, AI, EMI	Wind temperature, solar radiation, NO, NO_2, SO_2, O_3, and non-methane hydrocarbon concentration made at 5 sites Intensive period airborne sampling of traces	NO - 38 NO_2 - 49 O_3 - 133
[6] DOT (1979)	AIRSHED	Sacramento, CA	28 June, 1976 24 Aug., 1976	CARB, Cal. DOT Sacramento, Placer, Yolo County APCD SRAPC	Surface based hourly averaged NO, NO_2, O_3, CO, and non-methane hydrocarbon at 6 stations Surface wind, radiosonde, pebal, aircerft spirals	NO - 44 (6/28/76) ~ 46 (8/24/76) NO_2 - 44 (6/28/76) ~ 46 (8/24/76) O_3 - 42 (6/28/76) ~ 52 (8/24/76) CO - 44 (6/28/76) - 52 (8/24/76)
[7] DOT (1979)	AIRSHED	Denver, CO Denver, CO	29 July, 1975 28 July, 1976	APCO/CDOJE CDOHI	Surface based hourly averaged NO, NO_2, SO_2, O_3, non-methane hydrocarbon measurements at up to 11 sites	NO ~ 9 (7/29/75) ~ 9 (7/28/76) NO_2 - 18 (7/29/75) - 12 (7/28/76)

TABLE 6-4. SUMMARY OF AMS RECOMMENDED STATISTICS FOR ELSTAR

(1 hour averages)

Type of Performance Measure	General Type of Statistical Measure	Sample Size (n)	Chemical NO (pphm)	Species NO_2 (pphm)	O_3 (pphm)	CO (ppm)
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 9% confidence interval	132 [1] 11 [2]	$0.502 < 0.62 < 0.72$ [1]*	$0.35 < 0.49 < 0.61$ [1]	$0.78 < 0.84 < 0.88$ [1]	$0.42 < 0.55 < 0.66$ [1]
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} [SO^2/M + S_p^2/N]^{1/2}$ where $\alpha = 95$ percent confidence level		2.23 ± ? [1] 0.12 ± 3.06 [2]	2.70 ± ? [1] 4.5 ± 13.6 [2]	1.72 [1] 0.83 ± 11.8 [2]	0.20 [1] 1.8 ± 3.27 [2]
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi_{\nu\alpha} < \sigma_d^2 < S_d^2/\chi_{\nu 1-\alpha}$ where $\alpha = 95\%$ confidence level		$4.9 < 9.5 < 28$ [2]	$13.2 < 25.6 < 75$ [2]	$4.4 < 8.1 < 21$ [2]	$3.8 < 6.9 < 18$ [2]
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root mean square error $RMSE_d$ or the average absolute error $\lvert\bar{d}\rvert$		$RMSE_d = 3.1$ [2]	$RMSE_d = 6.6$ [2]	$RMSE_d = 2.8$ [2]	$RMSE_d = 3.1$ [2]

---

* Numbers in brackets refer to the citations presented in table 6.3

TABLE 6-5. A SUMMARY OF AMS RECOMMENDED STATISTICS FOR RPM

(Instantaneous Samples)

Type of Performance Measure	General Type of Statistical Measure	Sample Size (n)	Chemical NO (pphm)	Species NO_2 (pphm)	O_3 (pphm)
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	6 [3a] 18 [3b]	$0.66 < 0.86 < 0.95$ [3b]*	$0.64 < 0.85 < 0.94$ [3b]	$0.20 < 0.87 < 0.99$ [3a] $0.92 < 0.97 < 0.99$ [3b]
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} [SD^2/N + S_p^2/N]^{1/2}$ where α = 95% confidence level		2.23 ± ? [3a] 0.12 ± 3.06 [3b]	2.70 ± ? [3a] 4.5 ± 13.6 [3b]	1.72 [3a] 0.83 ± 11.8 [3b]
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi_{\nu\alpha} < \sigma_d^2 < S_d^2/\chi_{\nu}1-\alpha$ where α = 95% confidence level		$4.9 < 9.5 < 28$ [3b]	$13.2 < 25.6 < 75$ [3b]	$4.4 < 8.1 < 21$ [3b]
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root mean square error $RMSE_d$ or the average absolute error $\overline{\lvert d \rvert}$		$RMSE_d = 3.1$ [3b]	$RMSE_d = 6.6$ [3b]	$RMSE_d = 2.8$ [3b]

---

* Numbers in brackets refer to the citations presented in table 6-3.

TABLE 6-6. A SUMMARY OF AMS RECOMMENDED STATISTICS FOR THE AIRSHED MODEL

(1 hour averages)

Type of Performance Measure	General Type of Statistical Measure	Sample Size (n)	Chemical Species			
			NO (pphm)	NO$_2$ (pphm)	O$_3$ (pphm)	CO (pph)
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	22-80 [4] 38-133 [5] 42-52 [6] 9-80 [7] 17-137 [8] 203-310 [9] 184-331 [10]	0.60 < 0.75 < 0.85 [5] 0.37 < (0.57-0.71) < 0.82 [6] 0.29 < 0.75 < 0.93 [7] -0.40 < (-0.17-(-0.47)) < 0.65 [8] 0.33 < 0.41 < 0.48 [10]	0.15 < 0.25 < 0.34 [4] 0.04 < 0.27 < 0.48 [5] 0.25 < (0.47-0.56) < 0.71 [6] 0.04 < (0.37-0.65) < 0.83 [7] -0.44 < -0.05 < 0.357 [8] 0.55 < 0.61 < 0.66 [10]	0.76 < 0.83 < 0.84 [4] 0.72 < 0.78 < 0.83 [5] 0.60 < (0.74-0.84) < 0.90 [6] 0.57 < 0.70 < 0.79 [7] 0.81 < 0.86 < 0.90 [8] 0.48 < 0.57 < 0.65 [10]	-0.22 < (-0.13 - 0.68) < 0.73 [7] 0.33 < 0.41 < 0.48 [10]
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_\alpha[S_o^2/N + S_p^2/N]^{1/2}$ where α = 95% confidence level		0.17 ± ? [5] 2.5-3.6 ± ? [6]	0.06 ± ? [4] 0.54 ± ? [5] 1.4-2.0 ± ? [6] 0.33-2.8 ± ? [7] 0.25-0.70 ± ? [8] 1.56-2.0 ± ? [9] 0.08 ± ? [10]	-0.31 ± ? [8] 0.37 ± ? [5] 0.41-1.85 ± ? [6] 0.61-1.09 ± ? [7] 0.84-1.08 ± ? [8] -3.14-2.15 ± ? [9] -0.10 ± ? [10]	1.19-1.75 ± ? [6] 0.73-3.67 ± ? [7] 0.31 ± ? [10]
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu\, s_d^2/\chi_{\nu\alpha} < \sigma_d^2 < s_d^2/\chi_{\nu,1-\alpha}$ where α = 95% confidence level		0.38 < 0.43 < 0.49 [5] 1.35 < 1.88 < 2.90 [6] 1.1 < (2.0-8.4) < 22.9 [7] 1.6 < (1.2-6.0) < 3.7 [8] 1.5 < 1.7 < 1.9 [10]	1.22 < 1.39 < 1.6 [5] 0.96 < 1.3 < 1.9 [6] 0.4 < (0.7-1.75) < 4.8 [7] 0.81 < (1.3-2.25) < 4.3 [8] 10.9 < (12.4-16.1) < 18.5 [9] 0.46 < 0.52 < 0.59 [10]	5.8 < 6.6 < 7.6 [5] 1.7 < 2.1 < 2.6 [6] 2.5 < 3.3 < 4.6 [7] 1.4 < (1.7-2.0) < 2.5 [8] 40 < (47-66) < 56 [9] 0.23 < 0.27 < 0.32 [10]	0.52 < (0.71-0.84) < 1.25 [6] 1.0 < (1.9-2.2) < 4.2 [7] 0.41 < 0.47 < 0.54 [10]
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$		$RMSE_d$ = 0.46 [5] $RMSE_d$ = 8.13 - 14.84 [6] $RMSE_d$ = 8.25 - 16.25 [7] $RMSE_d$ = 2.65 - 15.0 [8] $RMSE_d$ = 1.7 [10]	$RMSE_d$ = 1.4 [5] $RMSE_d$ = 1.59 [6] $RMSE_d$ = 2.45 - 7.0 [7] $RMSE_d$ = 1.40 - 2.74 [8] $RMSE_d$ = 14.8 - 20.1 [9] $RMSE_d$ = 0.52 [10]	$RMSE_d$ = 6.6 [5] $RMSE_d$ = 2.2 - 5.52 [6] $RMSE_d$ = 3.6 - 4.48 [7] $RMSE_d$ = 2.4 - 3.16 [8] $RMSE_d$ = 52 - 76 [9] $RMSE_d$ = 0.28 [10]	$RMSE_d$ = 2.1 - 3.9 [6] $RMSE_d$ = 2.4 - 15.6 [7] $RMSE_d$ = 0.57 [10]
Spatial vector differences	Vector distance differences between the locations of $C_{o,n}$ and $C_{p,n}$ for the n-th highest observation		55 km and 400 (n=1) [5]	0 km and 0° (n=1) [5] 0 km and 0° (n=1) [8] 15 km and 0° (n=1) [7]	35 km and -50° (n=1) [5] 0 km and 0° (n=1) [6] 0 km and 0° (n=1) [7]	
Time differences	The difference between the occurrence of the n-th highest observed concentration and the n-th highest predicted concentration		0 hours (n=1) [4] 15-16 hours (n=1) [8] -25 hours (n=1) [10]	11 hours (n=1) [4] 9 hours (n-1) [5] 8-14 hours (n=1) [8] 0 hours (n=1) [10]	1 hour (n=1) [4] 3 hours (n=1) [5] -(2-3) hours (n=1) [8] 0 hours (n-1) [10]	-7 hours (n=1) [10]

7 MOBILE SOURCE MODELS

A group of air quality models have been constructed for or adapted to the specific problem of predicting concentrations near a roadway. Six such models are of particular interest in the present review. The names of these models as well as such pertinent information as the model developer and the date of development are given in table 7-1. Five of the models have been developed specifically for simulations of concentration of an inert pollutant near a roadway. The sixth model, PAL, can be used to model roadway line sources in addition to other categories of sources.

A survey of the available evaluation studies and their accompanying statistical measures is presented in table 7-2. Most models have at least two evaluation studies, with the exception of PAL, which has none. PAL, therefore, is dropped from further analysis in this section.

Some general comments can be made about the mobile source evaluation models. The evaluation studies again suffered a loss of information because means and standard deviations of predictions and observations were lacking, as were moment calculations of the residuals. Since raw data were rarely presented in the studies, the AMS recommended statistical measures could not be calculated. Other similarities among the studies include:

> The typical experiment site is a grade highway, with the monitors located within about 200 m of the highway. The species monitored were inert--either CO or SF_6.

> The predictions and observations were almost always matched by time and location.

> The model results were highly stratified, often being broken down by

  - Stability class

  - Receptor elevation

  - Wind direction

- Wind speed
- N highest observed or predicted
- Experiment location.

Many of the reported statistical measures came from a relatively high quality data base such as the GM sulfate experiment (Chock, 1977). The stratification problem was simplified, since the only common stratification among tests was by N highest and by experiment location. Only these two stratification distinctions are made in the present study.

The evaluation studies available on HIWAY-1 are described in table 7-3. Note the similarity of the tests and the statistical measures reported in table 7-4. The statistics for HIWAY-1 indicate overprediction and relatively high positive correlation. The cumulative frequency distributions, however, in all cases seem significantly dissimilar.

The newer and improved mobile source model, HIWAY-2, which supercedes HIWAY-1, has few evaluation studies (table 7-3). However, the studies are similar to those of HIWAY-1, and in one case, both models appear in the same study. Although fewer statistics are available for HIWAY-2 than for HIWAY-1, table 7-5 seems to indicate some improvement in HIWAY-2 over HIWAY-1. The cumulative frequency distributions agree, and the bias appears to have a mixed sign. The gross error is smaller than for the same evaluation study results for HIWAY-1.

The evaluation studies for CALINE3 are shown in table 7-3. Although the references are the same, the measurements on tracer experiments differed in location or tracer. The statistical results displayed in table 7-6 indicate a significantly high degree of positive correlation between observations and predictions. The model seems to show some indication of underprediction. However, the degree of underprediction seems small compared with the noise, and the confidence interval cannot be calculated in order to determine the bias' significance.

Descriptions of the GMLINE model evaluation studies are presented in table 7-3. The tracer experiments are similar to those presented for the three previous studies. Table 7-7 indicates that GMLINE has the largest positive correlations of any of the mobile source models. Its predicted and observed cumulative frequency distributions also seem to be identical.

The model APRAC-1A has the most unusual evaluation studies, as described in table 7-3. The model was tested in a street canyon siting in order to test its street canyon dispersion algorithms. The statistical measures, presented in table 7-8, indicate that significant positive correlation occurs, but to a substantially lesser degree than in the other models. The observed and predicted cumulative distributions also seem to be dissimilar.

In the course of compiling the AMS recommended statistics, several types of non-AMS statistics were encountered; the most prominent of these were

> Relative quantities created by forming a ratio of the observations, predictions, and residuals.

> Percentages of predictions within some set percentage of the observations.

Without the simple statistical moments of the observations or predictions, or without the original data, the information in these statistics cannot be properly interpreted in the framework of the AMS recommendations. However, as an additional aid to the reader, the major examples of such statistics were collected and are displayed in table 7-9 with appropriate qualifying statements.

TABLE 7-1. A SUMMARY OF BASIC INFORMATION ABOUT THE VARIOUS MOBILE SOURCE MODELS

Name	Developer and Reference	Description
HIWAY-1	J. R. Zimmerman and R. S. Thompson Environmental Protection Agency Research Triangle Park, North Carolina (Zimmerman and Thompson [1975])	Numerical integration of Gaussian point source algorithm
HIWAY-2	W. B. Petersen Environmental Protection Agency Research Triangle Park, North Carolina (Petersen [1980])	Numerical integration of Gaussian point source algorithm
CALINE 3	P.E. Benson State of California Department of Transportation Sacramento, California 95807 (Benson [1979])	Numerical integration of Gaussian point source algorithm
GMLINE	D. P. Chock General Motors Research Laboratories Warren, Michigan 48090 (Chock [1979])	Gaussian line source algorithm
APRAC-1A	R. L. Mancuso and F. L. Ludwig Stanford Research Institute Menlo Park, California (Mancuso and Ludwig [1972])	Downwind integrated line source model using Gaussian vertical dispersion and source-simplified box model for street canyons
PAL	W. B. Petersen Environmental Protection Agency Research Triangle Park, North Carolina (Petersen [1978])	Numerical integration of Gaussian point source algorithm

TABLE 7-2. A SUMMARY OF THE AVAILABILITY OF AMS STATISTICAL MEASURES FROM MODEL PERFORMANCE EVALUATION STUDIES OF MOBILE SOURCE MODELS

Model	Reference	Correlation Coefficients	Frequency Distribution Differences	Residual Measures			Space-Time Offsets	Graphical Measures
				Bias	Noise	Gross Error		
HIWAY-1	Maldonado and Bullin (1977)			X	X			
	Noll, Miller and Clagget (1978)	X		X				
	Rao and Kennan (1980)	X	X	X				CP
	Rao et. al (1980)	X	X					CP
	Rao and Visalli (1981)		X	X	X	X		CP
HIWAY-2	Rao and Visalli (1981)		X	X	X	X		CP
CALINE 3	Benson (1979)	X						SP
	Rao and Visalli (1981)		X	X	X	X		CP
GMLINE	Chock (1978)							
	Rao and Keenan (1980)	X	X	X				CP
	Rao et al. (1980)	X	X					CP
APRAC-1A	Johnson et al. (1973)	X						
	Dabberdt et. al. (1973)	X	X			X		CP
PAL	No appropriate studies submitted for review by developers							

TABLE 7-3. A DESCRIPTIVE SUMMARY OF EVALUATION STUDIES FOR THE MOBILE SOURCE MODELS

Reference	Model	Sample Size and Type	Specie	Averaging Period	Maximum Source Receptor Distance	Source Type	Number of Receptors	Land Use-Topography	Receptor Height
[1] Rao and Keenan (1980)	HIWAY-1 GMLINE	533; all 594; all	SF_6	1/2 hr	<100 m from roadway	Automobile	8	Flat highway track (rural)	4 - 100 m
[2] Rao and Keenan (1980)	HIWAY-1 GMLINE	133; all 133; all	SF_6	1 hr	< 100 m from roadway	Automobile	4	Interstate highway (rural)	4 - 30 m
[3] Rao and Visalli (1981)	HIWAY-1 CALINE3 HIWAY-2	594; all 595; all 594; all	SF_6	1/2 hr	< 100 m from roadway	Automobile	8	Flat highway track (rural)	4 - 100 m
[4] Rao and Visalli (1981)	HIWAY-1 CALINE3 HIWAY-2	51; top 10% 51; top 10% 59; top 10%	SF_6	1/2 hr	< 100 m from roadway	Automobile	8	Flat highway track (rural)	4 - 100 m
[5] Rao et al. (1980)	HIWAY-1 GMLINE	509; $\bar{U} > 1$ m/s 599; all	SF_6	1/2 hr	< 100 m from roadway	Automobile	8	Flat highway track (rural)	4 - 100 m
[6] Noll et al. (1978)	HIWAY-1	568; all	CO	1/4 hr	< 100 m from roadway	Automobile	5	Flat interstate highway (rural)	~ 10 m
[7] Maldonado and Bullin (1977)	HIWAY-1	1306; all	CO	1/4 hr	< 200 m from	Automobile roadway	Unknown	Flat but mixed-	3 - 10 m rural urban surroundings
[8] Benson (1979)	CALINE3	516; all	SF_6	1/2 hr	< 100 m from roadway	Automobile	8	Flat highway track (rural)	4 - 100 m
[9] Benson (1979)	CALINE3	266; all	CO	1 hr	< 50 m from roadway	Automobile	10	Flat interstate highway	1 - 13 m
[10] Benson (1979)	CALINE3	265; all	CO	1 hr	< 100 m from roadway	Automobile	15	Frequency in street canyon	1 - 20 m
[11] Benson (1979)	CALINE3	237; all	CO	1 hr	< 100 m from roadway	Automobile	15	Elevated freeway (urban)	1 - 20 m
[12] Dabberdt et al. (1973)	APRAC-1A	~ 670; all	CO	1 hr	< 125 m from nearest roadway	Automobile	4	Urban-street canyon	4 m
[13] Johnson et al. (1973)	APRAC-1A	~ 375; all	CO	1 hr	< 2 km intra urban area	Automobile	5	Urban-street canyon	2-30 m

TABLE 7-4. STATISTICAL MEASURES AND AGGREGATE STATEMENTS FOR HIWAY-1

(1/4 - 1 hour averages - ppb)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location) Time and Location	Time Only	Location Only	Unmatched	Aggregate Statement of Model Performance and Accuracy
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	533 [1] 133 [2] 594 [3] 51 [4] 509 [5] 568 [6] 1306 [7]	$0.75 < 0.78 < 0.81$ [1] $0.78 < 0.81 < 0.83$ [2] $0.78 < 0.81 < 0.83$ [5] $0.44 < 0.49 < 0.54$ [6]				A significant, positive relationship occurs between the observations and predictions at 95% level; the average correlation and its 95% confidence level is $0.68 < 0.72 < 0.75$.
Frequency distribution differences	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval					Fails the K-S test [1] Fails the K-S test [3] Fails the K-S test [5]	Observed and predicted cumulative frequency distributions are not identical at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} (S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		$-1.25 \pm ?$ [2] $-0.63 \pm ?$ [3] $-1.54 \pm ?$ [4] -0.3 ppm $\pm$ [7]				The average bias is -1.03 for SF_6, indicating that the model possibly overpredicts
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of confidence interval $\nu s_d^2/\chi^2_{\nu\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu 1-\alpha}$ where α = 95% confidence level		$0.89^2 < 0.93^2 < 0.98^2$ [3] $1.33^2 < 1.56^2 < 1.86^2$ [4] $3.45^2 < 3.56^2 < 3.68^2$ [7]				The average variance and its 95% confidence interval for SF_6 is $1.13^2 < 1.28^2 < 1.48^2$
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ or the average absolute error $\lvert\bar{d}\rvert$		$RMSE_d = 1.13$ [3] $RMSE_d = 2.18$ [4] $RMSE_d = 3.56$ ppm [7]				The average $RMSE_d$ and its range for SF_6 is $1.13 < 1.65 < 2.15$

N.B. Numbers in brackets refer to the citations presented in table 7.3.

TABLE 7-5. STATISTICAL MEASURES AND AGGREGATE STATEMENTS FOR HIWAY-2

(1/4 - 1-hour averages - ppb)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	594 [3] 51 [4]					
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval					Passes the K-S test [3]	In general, observations and predicted cumulative frequency distributions are identical at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha}(S_o^2/N + S_p^2/N)^{1/2}$ where $\alpha = 95\%$ confidence level		$-0.06 \pm ?$ [3] $0.37 \pm ?$ [4]				Inconclusive
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$, where $\alpha = 95\%$ confidence level		$0.34^2 < 0.36^2 < 0.37^2$ [3] $1.33^2 < 1.56^2 < 1.86^2$ [4]				The average variance and its 95% confidence interval is $0.97^2 < 1.13^2 < 1.34^2$
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ or the average absolute error $\lvert\overline{d}\rvert$		$RMSE_d = 0.36$ [3] $RMSE_d = 0.64$ [4]				The average $RMSE_d$ and its range is $0.36 < 0.50 < 0.64$

N.B. Numbers in brackets refer to the citations in table 7.3.

TABLE 7-6. STATISTICAL MEASURES AND AGGREGATE STATEMENTS FOR CALINE3
(1/4 - 1 hour averages - ppb)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	595 [3] 51 [4] 516 [8] 266 [9] 265 [10]	$0.76 < (0.79 - 0.93) < 0.98$ [8] $0.71 < (0.76 - 0.91) < 0.93$ [9] $0.90 < 0.92 < 0.93$ [10]				A significant, positive relationship occurs between the observations and predictions at the 95% level; the average correlation is 0.86
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 9% confidence interval					Passes the K-S test [3]	The observed and predicted cumulative frequency distributions are in general identical at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} (S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		$0.15 \pm ?$ [3] $0.75 \pm ?$ [4]				The average bias is a positive 0.45 which indicates a tendency toward model underprediction
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level		$0.37^2 < 0.40^2 < 0.42^2$ [3] $0.48^2 < 0.56^2 < 0.67^2$ [4]				The average variance and its 95% confidence interval is $0.43^2 < 0.48^2 < 0.56^2$
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ or the average absolute error $\lvert\bar{d}\rvert$		$RMSE_d = 0.43$ [3] $RMSE_d = 0.91$ [4]				The average $RMSE_d$ and its range is $0.43 < 0.67 < 0.91$

N.B. Numbers in brackets refer to the citations in table 7.3.

TABLE 7-7. STATISTICAL MEASURES AND AGGREGATE STATEMENTS FOR GMLINE

(1/4 - 1 hour averages - ppb)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	594 [1] 133 [2] 599 [5]	$0.90 < 0.91 < 0.92$ [1] $0.90 < 0.93 < 0.95$ [2] $0.86 < 0.88 < 0.90$ [3]				A significant, positive relationship occurs between the observations and predictions at the 95% level; average correlation and its 95% confidence interval is $0.89 < 0.90 < 0.92$
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval					Passes the K-S test [5]	The observed and predicted cumulative frequency distributions are identical at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha}(S_o^2/N + S_p^2/N)^{1/2}$ where $\alpha = 95\%$ confidence level		$0.20 \pm ?$ [2]				Inconclusive
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residuals as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$, where $\alpha = 95\%$ confidence level						
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ or the average absolute error $\lvert\bar{d}\rvert$						

N.B. Numbers in brackets refer to the citations in table 7.3.

TABLE 7-8. STATISTICAL MEASURES AND AGGREGATE PERFORMANCE STATEMENTS FOR APRAC-1A
(1 hour averages-ppm)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Association-correlation	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	~670 [12] ~375 [13]	0.13 < (0.20-0.40) < 0.46 [12] 0.53 < (0.60-0.70) < 0.75 [13]				A significant, positive relationship occurs between the observations and predictions at the 95% level; average correlation is 0.47
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval					Fails the K-S test [12]	The observed and predicted cumulative frequency distribution are in general not identical at the 95% level
Residual $(C_{obs} - C_{pred})$ differences (bias)	Average residual with confidence interval of $\bar{d} \pm t_{\alpha}(S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level						
Residual $(C_{obs} - C_{pred})$ differences (noise)	Variance of residuals as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$ where α = 95% confidence level						
Residual $(C_{obs} - C_{pred})$ differences (gross error)	Root-mean-square error $RMSE_d$ or the average absolute error $\overline{\|d\|}$		$RMSE_d$ = (2.6 - 3.9) [12]				The average RMSE and its range is 2.6 < 3.25 < 3.9

N.B. Numbers in brackets refer to the citations in table 7.3.

TABLE 7-9. SUMMARY OF THE NON AMS RECOMMENDED STATISTICS APPEARING IN THE MOBILE SOURCE MODEL EVALUATION STUDIES

Reference	Model(s)	Statistic Type
Rao and Visalli (1981)	HIWAY-1 HIWAY-2 CALINE3	● Percent of predictions with a factor of two of the matching observations ● Average percent error consisting as a ratio of the residuals over the observations
Rao et al. (1980)	GM HIWAY-1	● Linear regression coefficients such as the slope and intercept ● Standard deviation of the ratio of observations over predictions
Rao and Keenan (1980)	GM HIWAY-1	● Linear regression coefficient such as the slope and intercept ● Standard deviation of the ratio of observations over predictions ● The ratio of mean observed to mean predicted concentrations
Chock (1978)	GM	● Ratios of predicted to observed concentrations
Benson (1979)	CALINE3	● The ratio of residuals over the sum of the predicted and observed concentrations ● Linear regression coefficients such as the slope and intercept
Maldonado and Bullin (1977)	HIWAY-1	● Percent of predictions within some concentration range ● Linear regression coefficients such as the slope and intercept

8 COMPLEX TERRAIN MODELS

One of the most difficult tasks for air quality models is the simulation of pollutant dispersion in complex terrain. Complex terrain modifies the planetary boundary layer airflows, and complex, sometimes highly channeled, wind flows occur. In addition, the plume may be forced to impinge on elevated terrain under certain atmospheric conditions. Most of the Gaussian dispersion models such as CRSTER handle terrain only up to the height of the stack. This section focuses on models that can predict concentrations on terrain which varies in elevation over a height greater than the source height.

Six models possess varying degrees of ability to incorporate the effects of complex terrain:

> SHORTZ

> LONGZ

> COMPLEX I

> COMPLEX II

> IMPACT

> RTDM.WC

Of the six, the model IMPACT, as indicated by its description in table 8-1, is the most unlike the others. It is essentially a three-dimensional grid cell model that makes no steady-state or homogeneous assumptions about the dispersion conditions, as a Gaussian model does. In addition, three rural point source Gaussian models produced ground-level predictions when the terrain height exceeded the stack height:

> 3141/4141

> COMPTER

> PLUME5

These three Gaussian models are for the most part mere extensions of the CRSTER model that deal with the effects of complex terrain in a highly simplified and parameterized fashion. The major distinctions among the rest of the models are briefly summarized in table 8-1.

In general, most of the complex terrain models have been evaluated only once or twice. In some cases, the results presented are not model evaluations in the true sense of the word, but merely tests to see how well the complex terrain handling options work. Thus, the data bases available for evaluation are generally limited in size and in the range of dispersion conditions considered. A summary of the evaluation studies and the AMS recommended statistics associated with them is displayed in table 8-2. The most often reported statistics for the complex terrain models are the correlation coefficient and the bias. In order to fill out the survey, raw data presented in the evaluation reports of SHORTZ and RTDM.WC was used to compute statistics for model evaluation.

The results of the TACOMA-ASARCO modeling study were extracted from Bjorklund and Bowers (1979) and used to compute model evaluation statistics for SHORTZ. The descriptive summary in table 8-3 indicates that the modeling region size is unknown, since it was not presented in the study. The sample consisted of 26 hourly values for which the observations exceeded 0.25 ppm. The resulting statistics and aggregate statements are summarized in table 8-4. The results indicate

- A lack of bias
- Significant positive correlation
- Gross error measuring less than the observed mean
- Agreement between cumulative frequency distributions.

Despite the general indication of skill, the results are predicted on a limited time interval and sample size.

COMPLEX I and II model evaluation statistics were computed from the same data base, which is described in table 8-3. The release height varied during the study from 231 to 436 m. Only stable conditions were modeled. The results in table 8-5 indicate that, for the entire sample set, there is no significant positive correlation between the predictions and observations regardless of averaging period. However, the highest observed concentrations do seem to show significant positive correlation with their matching predictions. The table 8-6 results for COMPLEX II indicate less evidence of positive correlation.

The model IMPACT has been used extensively; however, few model evaluation statistics are available. Apparently this lack is due, in part, to the developer's opinion that normal means of model evaluation are of limited use because of data errors in model inputs. The evaluation statistics for the Cobb Valley-Northern California model simulations were found, and a description of the data base is given in table 8-3. The results, presented in table 8-7, indicate an extremely high degree of correlation. The scatter plot for the results indicates that most of the correlation is due to a few large concentrations which seem to be accurately predicted.

The model evaluation studies for RTDM.WC are described in table 8-3. In the first study only one receptor was used. Since sample size was small, data were taken from the scatter plot to calculate the AMS recommended statistics for this study. The authors of the second study warn against construing the results as a full-fledged model evaluation study. The statistical measures are presented in table 8-8. The results indicate

> A lack of significant positive correlation

> Presence of bias as model overprediction

> Failure of the cumulative frequency distributions to agree.

Table 8-3 indicates that the model 3141/4141 was evaluated in an industrialized area containing multiple point and area sources. Only the highest concentrations were present in the sample. The statistical results, shown in table 8-9, are somewhat inconclusive: the model seems to underpredict for 1-hour averages and overpredict for 24-hour averages. The confidence intervals are unknown; thus the significance of the biases is also unknown.

The model COMPTER was evaluated in a modeling region of unknown size. The descriptive summary of the evaluation study, displayed in table 8-3, indicates that both short- and long-term averages were taken. The short-term averages were very large samples. The model simulation with the options most suited for the modeling region were used in the evaluation study. The resulting statistical measures are presented in table 8-10. The correlations seem to be significant for all averaging periods, and the bias indicates a significantly nonzero underprediction. The cumulative frequency distributions were not tested, because of the lack of spatial representation that a single monitoring station has of the whole region.

TABLE 8-1. A SUMMARY OF BASIC INFORMATION ABOUT THE VARIOUS COMPLEX TERRAIN MODELS

Name	Developer	Description
SHORTZ	J. R. Bjorklund and J. F. Bowers H. E. Cramer Company, Inc. University of Utah Research Park Bjorklund and Bowers (1979)	A Gaussian dispersion model for multiple stack, building, and area sources. The model has algorithms for downwash, dry deposition, and Cramer dispersion coefficients for complex terrain
LONGZ	Same as SHORTZ	Same as SHORTZ except it runs in the long term (climatological) averaging mode where the meteorological input data is in the form of frequency distributions
COMPLEX I	Complex terrain team at Chicago Workshop on Air Quality Models, Chicago, Illinois (Turner [1980])	A Gaussian dispersion model with formulation similar to VALLEY. Special provision has been made for hourly concentration estimates. A buoyancy enhanced dispersion algorithm is included
COMPLEX II	Same as COMPLEX I	A Gaussian dispersion model with formulation similar to MPTER. A buoyancy enhanced dispersion algorithm as well as an additional high terrain handling algorithm is included
IMPACT	K. T. Tran and R. C. Sklarew Form and Substance, Inc. Westlake Village, California (Tran and Sklarew [1979])	A K-theory-grid cell model in three dimensions. It has a supporting diagnostic wind model called NEWEST. The model has chemistry and can handle both area and point sources in complex terrain
RTDM.WC	B. A. Egan, O. D'Errico, and Cosmo Vaudo Environmental Research and Technology Concord, Massachusetts (Morris [1980])	A Gaussian dispersion model that allows for direct plume impingement, terrain lifting of the plume, enhanced dispersion coefficients for complex terrain and buoyant plume entrainment
3141/4141	Enviroplan Staff Enviroplan, Inc. West Orange, N.J. (Enviroplan [1981])	Gaussian diffusion model for multiple co-located point sources. It is identical to CRSTER but handles complex terrain, buoyancy enhanced dispersion and power law correction to P-G dispersion coefficients for longer averaging time
COMPTER	Alabama Air Pollution Control Commission Montgomery, Alabama (AAPCC [1980])	Gaussian dispersion model for multiple point sources. Is similar to CRSTER except it has complex terrain handling options
PLUMES	J. Hsiung and H. Chase Pacific Gas and Electric Company San Francisco, California (Pacific Gas and Electric [1981])	Multiple point source Gaussian diffusion model. Buoyancy and directional wind shear enhanced dispersion are treated. Plume penetration through a stable layer plus simplified ozone chemistry are treated. A modification is made for terrain greater than stack height

TABLE 8-2. AVAILABLE AMS STATISTICAL MEASURES FROM MODEL PERFORMANCE EVALUATION STUDIES OF MODELS USED IN COMPLEX TERRAIN

Model	Reference	Correlation Coefficients	Frequency Distribution Differences	Residual Measures			Space-Time Offsets	Graphical Measures
				Bias	Noise	Gross Error		
SHORTZ*	Bjorklund and Bowers (1979)	X	X	X	X	X		SP, CP
COMPLEX I	NMEID (1981)	X	X					CP
COMPLEX II	NMEID (1981)	X	X					CP
IMPACT	Sklarew and Mirabella (1979)	X		X		X		SP
	Sklarew and Tran (1978)	X						
RTDM.WC*	Egan et al. (1979)	X	X	X	X	X		CP, SP
	Corkum and Bradstreet (1980)		X					CP
3141/4141	Ellis et al. (1980)			X				
COMPTER	AAPCC (1980)	X		X				CP
LONGZ	No applicable studies submitted for review by developers							
PLUME5	No applicable studies submitted for review by developers							

---

* Statistics calculated in this study from original prediction-observation data.

TABLE 8-3. A DESCRIPTIVE SUMMARY OF THE EVALUATION STUDIES OF THE COMPLEX TERRAIN MODELS

Reference	Model	Species	Topography	Land Use	Surface Cover	No. of Receptors	Minimum Source Receptor Distance (km)	Maximum Source Receptor Distance (km)	Source Type	Averaging Time	Size-Type
[1] Bjorklund and Bowers (1979)	SHORTZ	SO_2	Complex	Rural-urban	Mixed houses-trees	7	Unknown	Unknown	Buoyant smelter stack	1 hour	26; highest
[2] NMEID (1981)	COMPLEX I	SF_6	Complex	Rural	Desert	23	Unknown	Unknown	Tethered balloon 231 436 m	1-3 hr	1099; all 61-65; highest 630; all 3 hr
[3] NMEID (1981)	COMPLEX II	SF_6	Complex	Rural	Desert	23	Unknown	Unknown	Tethered balloon 231-436 m	1-3 hour	1099; all (61-65); highest 630; all 3 hr
[4] Sklarew and Mirabella (1979) and Sklarew and Tran (1978)	IMPACT	SF_6	Complex	Rural	Trees	6	Unknown	30	Tethered balloon release	1 hr	285; all
[5] Egan, D'Errico and Vaudo (1979)	RTDM.WC	SO_2	Complex	Rural	Unknown	1	3.0	3.0	1 stack 178 m	1 hr	38 hrs; all
[6] Corkum and Bradstreet (1980)	RTDM.WC	SO_2	Complex	Rural	Trees	2	0.5	0.3	4 stacks 23-29 m	24 hrs	41; all
[7] Ellis, et al. (1980)	3141/4141	SO_2	Complex	Moderately industrialized	Unknown	7	0.05	30	Multiple point and area sources	3, 24 hr	112; highest
[8] AAPCC (1980)	COMPTER	SO_2	Complex	Rural	Trees-fields	8	Unknown	Unknown	8 stacks 82-152 m	1, 3, 24 hr annual	70; all 8; all

TABLE 8-4. STATISTICAL MEASURES AND AGGREGATE STATEMENTS FOR SHORTZ

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z Statistic for the 95% confidence interval	26 [1]	$0.23 < 0.58 < 0.79$ [1]				A significant, positive relationship occurs between the observations and predictions at the 95% level; see the statistics
Frequency distribution differences	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval					Passes the K-S test [1]	The observed and predicted cumulative frequency distributions are identical at the 95% level
Residual $(C_{obs} - C_{pred})$ differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} (S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		0.022 ± 0.064 [1]				There is no significant bias at the 95% level
Residual $(C_{obs} - C_{pred})$ differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level		$0.17^2 < 0.21^2 < 0.27^2$ [1]				None; see the statistic
Residual $(C_{obs} - C_{pred})$ differences (gross error)	Root-mean-square error $RMSE_d$ and the average absolute error $\overline{\lvert d \rvert}$		$RMSE_d = 0.21$ [1] $\overline{\lvert d \rvert} = 0.16$ [1]				None; see the statistic

---

N.B. Numbers in brackets refer to the citations presented in table 8-13.

TABLE 8-5. STATISTICAL MEASURES AND AGGREGATE STATEMENTS FOR COMPLEX I

(1-3 Hour Average)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	1099 [2a] 65 [2b] 630 [2c] 61 [2d]	$-0.06 < 0.00 < 0.06$ [2a] $0.09 < (0.33\text{-}0.38) < 0.57$ [2b] $-0.07 < (0.10\text{-}0.03) < 0.11$ [2c] $0.09 < (0.33\text{-}0.38) < 0.58$ [2d]				The full sample correlations are not significant at the 95% level. The sample of the highest concentrations indicate a significant positive correlation. The average correlation is 0.36
Frequency distribution differences	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval						
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} (S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level						
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $v s_d^2/\chi^2_{v,\alpha} < \sigma_d^2 < v s_d^2/\chi^2_{v,1-\alpha}$, where α = 95% confidence level						
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ and the average absolute error $\lvert\bar{d}\rvert$						

N.B. Numbers in brackets refer to the citations presented in table 8-3.

(2a) 1 hour average

(2b) 1 hour average-highest only

(2c) 3 hour average

(2d) 3 hour average-highest only

TABLE 8-6. STATISTICAL MEASURES AND AGGREGATE STATEMENTS FOR COMPLEX II

(1-3 Hour Average)

			Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				
Performance Measure	General Statistical Measure	Sample Size (n)	Time and Location	Time Only	Location Only	Unmatched	Aggregate Statement of Model Performance and Accuracy
Associative-	Pearson moment correlation coefficient	1099 [3a]	$-0.08 < -0.02 < 0.04$ [3a]	$-0.15 < (0.10\text{-}0.13) < 0.36$ [3b]			There does not seem to be a significant positive correlation between the predictions and observations at the 95% level
[illegible]	[illegible]	45 [3b]	$0.12 < (0.13\text{-}0.18) < 0.41$ [3b]	$0.23 < (0.46\text{-}0.52) < 0.68$ [3d]			
	confidence interval	630 [3c]	$-0.13 < -0.05 < 0.03$ [3c]				
		61 [3d]	$-0.13 < (0.13\text{-}0.18) < 0.41$ [3d]				
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval						
Residual $(C_{obs} - C_{pred})$ differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} (S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level						
Residual $(C_{obs} - C_{pred})$ differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $vS_d^2/\chi^2_{v,\alpha} < \sigma_d^2 < vS_d^2/\chi^2_{v,1-\alpha}$, where α = 95% confidence level						
Residual $(C_{obs} - C_{pred})$ differences (gross error)	Root-mean-square error $RMSE_d$ and the average absolute error $\overline{\lvert d \rvert}$						

N.B. Numbers in brackets refer to the citations presented in table 8-3.

(3a) 1 hour average

(3b) 1 hour average-highest only

(3c) 3 hour average

(3d) 3 hour average-highest only

TABLE 8-7. STATISTICAL MEASURES AND AGGREGATE STATEMENTS FOR IMPACT
(ppt--1-Hour Average)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	285 [4]	$0.89 < 0.91 < 0.92$ [4]				The observations and predictions seem to be significantly, positively correlated at the 95% level
Frequency distribution differences	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval						
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} (S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		$-0.09 \pm ?$ [4]				Inconclusive
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level						
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error RMSE_d or the average absolute error $\lvert\bar{d}\rvert$		$\lvert\bar{d}\rvert = 0.89$ [4]				None; see statistic

---

N.B. Numbers in brackets refer to the citations presented in table 8-3.

TABLE 8-8. STATISTICAL MEASURES AND AGGREGATE STATEMENTS FOR RTDM.WC

(ppm--1-24 Hour Averages)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	38 [5] 41 [6]	$-0.17 < 0.16 < 0.46$ [5]				There is not a significant positive confidence relationship between the observation and predictions at the 95% level
Frequency distribution differences	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval					Fails K-S test [5] Fails K-S test [6]	The observed and predicted cumulative frequency distributions are not identical at the 95% level
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha}(S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		-0.28 ± 0.6 [5]				There is an average bias at the 95% level which indicates general overprediction by the model
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level		$3.0^2 < 3.6^2 < 4.4^2$ [5]				None; see the statistic;
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ and the average absolute error $\lvert\bar{d}\rvert$		$\lvert\bar{d}\rvert = 2.8$ [5] $RMSE_d = 4.5$ [5]				None; see statistic

---

N.B. Numbers in brackets refer to the citations presented in table 8-3.

TABLE 8-9. STATISTICAL MEASURES AND AGGREGATE STATEMENTS FOR 3141/4141
(3-24 Hour Averages-$\mu g/m^3$)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	122 [7]					
Frequency distribution differences	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval						
Residual $(C_{obs} - C_{pred})$ differences (bias)	Average residual with confidence interval of $\pm t_{\alpha}(S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level				734 ± ? [7] -5 ± ? [7]		Inconclusive
Residual $(C_{obs} - C_{pred})$ differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level						
Residual $(C_{obs} - C_{pred})$ differences (gross error)	Root-mean-square error $RMSE_d$ and the average absolute error $\overline{\lvert d \rvert}$						

N.B. Numbers in brackets refer to the citations presented in table 8-3.

TABLE 8-10. STATISTICAL MEASURES AND AGGREGATE PERFORMANCE STATEMENTS FOR COMPTER ($\mu g/m^3$)

(a) 1, 3, 24 Hour Averages

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location): Time and Location	Time Only	Location Only	Unmatched	Aggregate Statement of Model Performance and Accuracy
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	70 [8a] 70 [8b] 70 [8c]		$0.67 < 0.78 < 0.86$ [8a] $0.75 < 0.84 < 0.90$ [8b] $0.84 < 0.90 < 0.94$ [8c]			A significant, positive relationship occurs between the observations and predictions at the 95% level; the average correlation and its 95% confidence level is $0.75 < 0.84 < 0.90$
Frequency distribution differences	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval						
Residual $(C_{obs} - C_{pred})$ differences (bias)	Average residual with confidence interval of $\pm t_\alpha (S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		12 ± 7 [8a] reported but in error [8b] reported but in error [8c]				There is an average bias significant at the 95% level. The bias indicates model underprediction
Residual $(C_{obs} - C_{pred})$ differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$ where α = 95% confidence level						
Residual $(C_{obs} - C_{pred})$ differences (gross error)	Root-mean-square error $RMSE_d$ and the average absolute error $\lvert\bar{d}\rvert$						

N.B. Numbers in brackets refer to the citations presented in table 8-3.

[8a] 1 hour average

[8b] 3 hour average

[8c] 24 hour average

TABLE 8-10 (Concluded)

($\mu g/m^3$)

(b) Annual Average

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient with Z statistic for the 95% confidence interval	8 [8]		0.30 < 0.83 < 0.97 [8]			A significant, positive correlation at the 95% level; see statistic
Frequency distribution differences	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval						
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} (S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% conficence level		-43 ± ? [8]				There seems to be a general indication of model over-prediction
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residual as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level						
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ and the average absolute error $\lvert\overline{d}\rvert$						

N.B. Numbers in brackets refer to the citations presented in table 8-3.

9 PLUME VISIBILITY MODELS

Visibility models are rather complex since they must not only predict concentrations, as the foregoing models do, but also several measures of what an observer calls visibility. This section is concerned with the performance and accuracy of two plume visibility models, namely

> PLUVUE

- A plume visibility model developed by Systems Applications, Inc. under contract with the EPA (Johnson et al. [1980]).

> VISIBILITY

- A plume visibility model developed by Environmental Research and Technology, Inc. (Drivas et al. [1978])

A description of PLUVUE can be found in Johnson et al. (1980); ERT's VISIBILITY has been described by Drivas et al. (1980).

Both plume visibility models consist of several modules. Plume visibility models predict up to as many as nine measures of concentration or visibility. Both models have been developed in a similar fashion and contain

> Plume rise and dispersion algorithms

> Plume chemistry algorithms

> Atmospheric optics algorithms.

A number of aerometric parameters need to be accurately predicted in order to arrive at a description of atmospheric visibility. Several parameters, presented in the evaluation study of Bergstrom et al. (1981), are also presented in this study and are listed below.

Parameter	Description or Definition
NO_2/NO_x ratio	Measured at plume centerline
Integrated plume NO_2	Defined as ${}_{o}^{L}\,[NO_2]\,dx$ where L = length of sight path x = distance along site path.
Integrated plume b_{sp} (450 nm)	Defined as ${}_{o}^{L}\,b_{sp}\,(450\ nm)\,dx$ where b_{sp} (450 nm) is the particualte scattering coefficient at 450 nm
Plume contrast , $C(\lambda)$	Defined as $C(\lambda) = \frac{I_p(\lambda)}{I_s(\lambda)} - 1$ λ = wavelength where $I_p(\lambda)$ = light intensity of the plume $I_s(\lambda)$ = light intensity of the sky this parameter is a measure of what the human eye perceives
Blue-red ratios	Defined as $\rho = \frac{I_p(450\ nm) \cdot I_s(630\ nm)}{I_p(630\ nm) \cdot I_s(450\ nm)}$ This parameter is a measure of plume discoloration

The data base used for the evaluation of both visibility models was selected from the VISTTA* 1979 field programs.

The June-July and December 1979 EPA-VISTTA field programs consisted of the recording of a set of coordinated measurements to characterize the impact of a specific power plant--the Navajo Generating Station--located in the vicinity of mandatory class I areas. Measurements of the emission rates from the stacks and of the attendant meteorological conditions were carried out. Telephotometers were used to measure the visual effects of the plume, and plume concentration measurements were performed with an instrumented airplane. The aircraft made continuous measurements of SO_2, NO_x, and O_3 concentrations and aerosol-scattering coefficients. To the extent possible, the flight paths were aligned with the telephotometer sight paths so that concentration measurements throughout the plume were made at the plume downwind distance corresponding to the telephotometer measurements.

Twenty case studies were selected from the 1979 data base, including,

> 2 cases from July 13

> 14 cases from December 7

> 4 cases from December 15

These cases have been described in some detail by Bergstrom et al. (1981). The VISTTA data base provides a good evaluation study for power plants which control the emission of aerosols. The evaluation of the visibility models with high aerosol concentration plumes remains to be done.

In all cases the observations and predictions are matched by time and location as closely as possible. Four statistical measures recommended by the AMS Workshop were computed for each predicted variable and for each model. The statistical measures are presented in tables 9.1 and 9.2 for PLUVUE and ERT VISIBILITY, respectively.

One of the general observations that can be made about the statistical summary in tables 9.1 and 9.2 is that both models tend to have very similar values for each of the validation variables. A natural division point in the tables can be made between concentration and optical measures. The concentration measures, which are in the first three rows in tables 9.1 and 9.2, indicate that

---

* Visibility Impairment due to Sulfur Transport and Transformation in the Atmosphere.

> Of the three concentration measures only the integrated plume NO_2 concentration shows a significant positive correlation at the 95 percent level for either plume model.

> The bias is not significant at the 95 percent level for either of the three concentration measures, or for either model.

> The measure of noise is generally much larger than the square of the bias.

The optical measures, which are in the last six rows of tables 9.1 and 9.2, generally show that for either model

> The integrated b_{sp} (450 nm) has a significant positive correlation at the 95 percent level.

> Only the plume contrast at 550 nm has a significant positive correlation at the 95 percent level.

> The blue-red ratios do not have significant positive correlation at the 95 percent level.

For the measures of plume contrast, PLUVUE seems to have a significant positive bias at the 95 percent level, which indicates underprediction. The ERT VISIBILITY model also has a significant bias; however, it is negative in sign and indicates overprediction. The noise of ERT VISIBILITY predictions tends to be much larger than PLUVUE's for most of the plume contrast measures. The ERT VISIBILITY blue-red ratio predictions do not appear to have a significant bias, whereas those of PLUVUE have a significantly positive bias, indicating model underprediction.

TABLE 9-1. STATISTICAL MEASURES AND AGGREGATE PERFORMANCE STATEMENTS FOR PLUVUE

Variable	Pearson Moment Correlation Coefficient with Z Statistic--95% Confidence Level	Average Residual with T Statistic--95% Confidence Interval	Residual Variance with chi square--95% Confidence Interval	Root-Mean Square Error
NO_x maximum plume concentration (ppm)	$-0.09 < 0.37 < 0.07$	-0.14 ± 0.29	$0.21 < 0.33 < 0.64$	0.57
NO_2/NO_x at plume centerline	$-0.28 < 0.19 < 0.58$	-0.02 ± 0.05	$0.007 < 0.012 < 0.022$	0.11
Integrated plume NO_2 concentration (ppm - km)	$-0.05 < 0.40 < 0.72$	-0.04 ± 0.07	$0.012 < 0.02 < 0.04$	0.14
Integrated plume b_{sp} (450 nm)	$0.50 < 0.77 < 0.90$	0.005 ± 0.17	$0.06 < 0.09 < 0.18$	0.30
Plume contrast at 405 nm	$-0.63 < -0.26 < 0.21$	0.04 ± 0.026	$0.0036 < 0.0057 < 0.011$	0.08
Plume contrast at 450 nm	$-0.56 < -0.15 < 0.31$	0.07 ± 0.025	$0.003 < 0.005 < 0.009$	0.10
Plume contrast at 550 nm	$0.17 < 0.57 < 0.81$	0.05 ± 0.016	$0.0005 < 0.0008 < 0.0016$	0.056
Plume contrast at 630 nm	$-0.10 < 0.36 < 0.69$	0.05 ± 0.014	$0.0005 < 0.0008 < 0.0016$	0.058
Blue-red ratios	$-0.68 < -0.34 < 0.12$	0.07 ± 0.03	$0.0055 < 0.0087 < 0.016$	0.116

TABLE 9-2. STATISTICAL MEASURES AND AGGREGATE PERFORMANCE STATEMENTS FOR ERT VISIBILITY

Variable	Pearson Moment Correlation Coefficient with Z Statistic--95% Confidence Level	Average Residual with T Statistic--95% Confidence Interval	Residual Variance with chi square--95% Confidence Interval	Root-Mean Square Error
NO_x maximum plume concentration (ppm)	$-0.13 < 0.33 < 0.67$	-0.30 ± 0.42	$0.51 < 0.82 < 1.57$	0.93
NO_2/NO_x at plume centerline	$-0.30 < 0.16 < 0.56$	0.044 ± 0.052	$0.009 < 0.014 < 0.027$	0.12
Integrated plume NO_2 concentration (ppm - km)	$-0.04 < 0.41 < 0.72$	0.042 ± 0.066	$0.0116 < 0.019 < 0.037$	0.14
Integrated plume b_{sp} (450 nm)	$0.52 < 0.78 < 0.91$	-0.07 ± 0.17	$0.056 < 0.089 < 0.17$	0.30
Plume contrast at 405 nm	$-0.39 < 0.066 < 0.49$	-0.65 ± 0.022	$0.002 < 0.003 < 0.006$	0.084
Plume contrast at 450 nm	$-0.36 < 0.099 < 0.52$	-0.047 ± 0.021	$0.0017 < 0.0028 < 0.0052$	0.070
Plume contrast at 550 nm	$0.20 < 0.59 < 0.82$	-0.035 ± 0.032	$0.0024 < 0.0039 < 0.0073$	0.070
Plume contrast at 630nm	$-0.23 < 0.24 < 0.62$	-0.078 ± 0.067	$0.02 < 0.032 < 0.06$	0.19
Blue-red ratios	$-0.23 < 0.24 < 0.62$	0.006 ± 0.041	$0.0087 < 0.014 < 0.026$	0.11

10 LONG-RANGE TRANSPORT MODELS

Many air quality and environmental impacts occur on the regional scale. In order to determine cause and effect relationships between the emissions and the impacted regions, long-range transport models have been developed. Although the scale of applicability ranges from the city-scale to continental scales, the models typically treat such processes as long-range transport and diffusion, chemical transformations, and dry and wet pollutant deposition. Presently most long-range transport models are used for SO_2/SO_4 impact assessments. However, several models have been developed to treat NO_x chemistry and photochemical smog formation.

Several long-range transport models developed recently include

- MESOPUFF.
- MESOPLUME.
- MESOGRID.
- RTM (Regional Transport Model).
- MTDDIS (Mesoscale Transport, Diffusion and Deposition Model for Industrial Sources.
- UWATM-SO_x.
- RADM (Random-Walk Advection and Dispersion Model).

Although these models were submitted for model evaluation reviews, only two of the models (UWATM-SO_x and RADM) had evaluation efforts documented. Of these two, only one (RADM) presented any evaluation statistics. Although UWATM-SO_x was used to predict wet deposition of sulfate at several sites, no statistics were presented; nor were there sufficient data to generate statistics. A third model, RTM, was in the process of being evaluated when the review was conducted. Although not documented at this time, the evaluation statistics have been included in this review.

A listing of model developers and an overview of model attributes for the three models are presented in table 10-1. The three models treat only sulfur chemistry, and all contain dry deposition parameterization. Two of the three parameterize precipitation-scavenging. Two of the models are grid-type, and the third is a Lagrangian particle-type model.

A summary of the model performance studies and the statistics generated for the RTM and RADM are shown in table 10-2. The unpublished evaluation effort for RTM is currently being undertaken by Systems Applications, Inc. The results are expected to be available in 1982 (Liu, private communication).

The data base used to evaluate the RTM is summarized in table 10-3. During July 1978, a high sulfate episode occurred over portions of the eastern United States. The episode date coincided with an intensive observation period in the SURE program; therefore, data capture was quite high. The RTM was exercised on an 80 x 80 km grid, and the concentrations of SO_2 and $SO_4^=$ were stored on a 3-hour basis for later analysis. Statistical performance measures for SO_2 and $SO_4^=$ are presented in table 10-4. Both SO_2 and $SO_4^=$ concentrations represent 24-hour averages.

The statistical measures in table 10-4 indicate that whereas a significant positive correlation occurs for SO_2 for some location-time matches of predictions and observations, others do not show significant correlation. There seems to be no systematic reason for such varying results, and consequently the aggregate statements are inconclusive. The SO_2 bias also seems to have no systematic behavior, either in sign and degree or in significance. Thus, the aggregate statement for bias is also labeled inconclusive.

The sulfate, $SO_4^=$, model evaluation statistics indicate a generally positive significant correlation at the 95 percent confidence interval. The exceptions are for small sample sizes. The bias statistics are consistently positive and significant at the 95 percent level, indicating model underprediction. The gross error seems to consist primarily of the average residual (bias), with the noise being much smaller.

Two independent evaluations of RADM were documented in Goodin, et al. (1980). A summary of the data bases utilized in these studies is presented in table 10-3. During the San Francisco Bay Area evaluation, the model was exercised for four 2-day scenarios. Sulfate predictions (24-hour averages) were compared with measurements for each of the episodes. Statistical measures reported in Goodin et al. (1980) express average performance over the four scenarios; these statistics are reported in table 10-5. Although sample size was reported (61 total data points), information on predicted and measured sulfate concentration variances and on residual variance was absent, thus precluding a calculation of confidence intervals. Additional statistics presented in Goodin et al. (1980), but omitted from table 10-5, include percentages of residuals less than 1 and 2 $\mu g\ m^{-3}$, and scatter plots of measured and predicted sulfate concentrations for all modeled scenarios.

A second RADM evaluation study was performed for hourly SO_2 concentrations only. Although statistics were presented in the documentation, a complete set of raw data was available for two surface stations and one elevated monitor covering 5 simulation days. A scatter plot of observed and predicted SO_2 concentrations is also presented in Goodin et al. (1980). Shown in table 10-5 are the AMS recommended statistics for hourly SO_2 residuals generated from these raw data. The statistics indicate that the predictions and observations are positively correlated at the 95 percent level. The SO_2 does not seem to have a significant bias. Furthermore, the gross error seems to consist primarily of noise and not bias.

TABLE 10-1 A SUMMARY OF BASIC INFORMATION ABOUT THE VARIOUS LONG-RANGE TRANSPORT MODELS

Name	Developer	Model Description
RTM	Systems Applications, Inc. Liu et al. (1981)	• $1\frac{1}{2}$ layer grid model (surface and mixed layer) • Gaussian plume (bi-variate normal) submodel for grid initialization • SHASTA scheme for numerical integration • Linear SO_2/SO_4 transformation (diurnal variation) • Dry deposition in a function of flux through surface layer and absorbtion at ground (vegetation dependent) • Wet deposition-SO_4, using Scott (1978) parameterization; workout and rainout from three cloud types • Wet deposition-SO_2, reversible using equilibrium parameterization of Hales and Sutter (1973) • Atmospheric optics using Kochsmeider relationship • Objectively analysed wind field, precipitation field • Spatially and temporally varying exposure class • Plume rise algorithm
UWATM-SOX	University of Wisconsin Ragland and Wilhening (1980)	• 2-dimensional Eulerian grid model • First-order fully implicit numerical scheme • Plume rise calculation for major point sources • Boundary layer characterized by wind and diffusivity profile dependent on micrometeorological variables • Linear SO_2/SO_4 transformation (daytime value and nighttime value) • Variable deposition velocity over region dependent on atmospheric and surface resistance • Scotti (1978) SO_4 wet removal assuming Bergeron-type clouds • Equilibrium SO_2 scavenging, different for rain and snow • Hourly gridded meteorological data • Estimated precipitation pH by SO_4 only
RADM model	Dames and Moore Runchal (1980) Goodin et al. (1980) Austin et al. (1981)	• Lagrangian random-walk advection and diffusion (deterministic advection, random dispersion • Dry deposition velocities for SO_2 and SO_4 • Motion of particulates is traced over a period of time • Diffusion is based on probability distribution of random motion, but normally a diffusivity (constant) or tensor is input to scale the probability distribution • First-order chemical reaction or radioactive decoy allowed
MESOPUFF	Environmental Research and Technology, Inc. Benkley and Bass (1979a)	• Lagrangian puff model • Vertical concentration distributions, either Gaussian or uniform, depending on user specification • Horizontal distribution of circular puff is assumed to be Gaussian • Dispersion parameters determined by Turner formulas for downwind distances < 100 km and by Heffter formulas for distance >100 km • Plume rise algorithm • Dry deposition of SO_2 and $SO_4^=$ via deposition velocity specifications • Linear $SO_2 \rightarrow SO_4$ chemical transformation • Objectively determined wind fields, mixing depths, and stabilities from preprocessor
MESOPLUME	Environmental Research and Technology, Inc. Benkley and Bass (1979b)	• Lagrangian plume-segment model • Vertical concentration distribution either Gaussian or uniform depending on user specification • Horizontal distribution of plume segment is assumed to be Gaussian • Dispersion parameters determined by Turner formulas for downwind distances < 100 km and by Hefter formulas for distance > 100 km • Plume rise algorithm • Dry deposition of SO_2 and $SO_4^=$ via deposition velocity specification • Linear $SO_2 \rightarrow SO_4$ chemical transformation • Objectively determined wind fields, mixing depths, and stabilities from preprocessor
MESOGRID	Environmental Research and Technology, Inc. Morris et al. (1979)	• 3-dimensional Eulerian grid model • "Method of Moments" advection scheme • Variable K_z profile • Plume rise algorithm • Linear $SO_2 \rightarrow SO_4$ transformation rate • Dry deposition of SO_2 and SO_4 via deposition velocity specification • Objectively determined wind fields, mixing depths, and stabilities from preprocessor
MTDDIS	Rockwell International Wang et al. (1980)	• Lagrangian puff model • Hourly surface observation used with 12-hour upper air observation to specify hourly stability winds, mixing heights, plume rise heights • Gaussian distribution in horizontal $\sigma_y = 0.5t$ • Gaussian distribution in vertical (multiple reflection) with $\sigma_z = (2K_z t)^{1/2}$ • Pollutant half-life specification • Deposition velocity concept spatially and temporally uniform • Washout ratio for SO_4 and SO_2 precipitation is rate-dependent

Reproduced from best available copy.

TABLE 10-2. A SUMMARY OF THE AVAILABILITY OF AMS STATISTICAL MEASURES FROM MODEL PERFORMANCE EVALUATION STUDIES OF LONG-RANGE TRANSPORT MODELS

Model	Reference	Correlation Coefficients	Frequency Distribution Differences	Residual Measures			Space Time Offsets	Graphical Measure
				Bias	Noise	Gross Error		
RTM	Stewart, private communication	x		x	x	x		
RADM	Goodin et al. (1980)	x		x		x		SP
MESOPUFF	No appropriate studies submitted for review by developers							
MESOPLUME	No appropriate studies submitted for review by developers							
MESOGRID	No appropriate studies submitted for review by developers							
MTDDIS	No appropriate studies submitted for review by developers							
UWATM-SO_x	No appropriate studies submitted for review by developers							

TABLE 10-3 A DESCRIPTIVE SUMMARY OF DATA BASES USED IN THE LONG RANGE MODELS

Reference	Model	Location	Date(s)	Program Participants*	Data Base/Measurements Made	Sample Size for Model Evaluation
[1] Stewart, personal communication	RTM	Eastern United States	17-24 July 1978	EPRI, ERT, GCA, MRI, RTI	National Climatic Center radiosonde data Surface weather observation EPRI/SURE data base including emission inventory and 9 class I stations measuring hourly average SO_2, NO, NO_x, O_3, temperature, data point, 24-hour Hivol aerosol sampler; 45 ground-level Class II stations measuring SO_2 and 24-hour Hivol aerosol sampler Limited craft data	307; all SO_2 8; time matched SO_2 47; location matched SO_2 372; all SO_4 8; time matched SO_4 50; location matched SO_4
[2] Goodin et al. (1980)	RADM	San Francisco Bay Area/Carquinez Straits	14-16 June 1975 24-26 Dec. 1976 11-12 June 1977 20-21 Dec. 1977	BAAQMD, NASN, ARB	Hourly averaged wind speeds and direction from BAAQMD network; winds, temperature aloft at Oakland airport; total evaluation, visibility at various airports Valid SO_2 and SO measurements at less than 20 sites Emissions from BAAQMD	13; (Jan/75) 16; (Dec./76) 19; (June/77) 13; (Dec./77)
[3] Goodin et al. (1980)	RADM	Port Valdez, AL	9-10 May 1978 23-24 Sept. 1978 29 June 1978	APSC, ADEC	SO_2 measurements at 5 locations Surface wind speed and direction at 9 stations Cloud cover, veiling height at 2 locations Emissions from APSC Occasional acoustic sounder data. Elevated SO_2 monitor	175; all

---

* EPRI-Electric Power Research Institute
ERT-Environmental Research and Technology, Incorporated
GCA-GCA Corporation
MRI-Meteorology Research Incorporated
RTI-Research Triangle Institute
ADEC - Alaska Department of Environmental Conservation
APSC - Alaska Pipeline Service Company
ARB - Air Resources Board
BAAQMD - Bay Area Air Quality Management District
NASN - National Air Surveillance Network

TABLE 10-4. STATISTICAL MEASURES AND AGGREGATE STATEMENTS FOR RIM

(24-Hour Average SO_4 $\mu g/m^3$)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location)				Aggregate Statement of Model Performance and Accuracy
			Time and Location	Time Only	Location Only	Unmatched	
Associative-correlative	Pearson moment correlation coefficient r with Z statistic for the 95% confidence interval	372 [1a] 8 [1b] 50 [1c] 186 [1d] 4 [1e]	$0.17 < 0.27 < 0.36$ [1a] $0.06 < 0.20 < 0.33$ [1d]	$0.40 < 0.61 < 0.76$ [1c] $-0.05 < 0.35 < 0.66$ [1f]	$-0.42 < 0.40 < 0.86$ [1b] $-0.92 < 0.35 < 0.98$ [1e]		Inconclusive
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval	25 [1f] 37 [1g]					
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha}(S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		-18.2 ± 5.7 [1a] -11.3 ± 10.3 [1d]	$+118.8 \pm 154.4$ [1c] $+259.5 \pm 237.9$ [1f]	6.2 ± 25.1 [1b] 22.7 ± 50.0 [1e]	-27.9 ± 5.6 [1a] -31.4 ± 10.0 [1d] -8.6 ± 44.0 [1g]	Inconclusive
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residuals as an estimator of σ_d^2 confidence interval: $\nu S_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu S_d^2/\chi^2_{\nu,1-\alpha}$ where α = 95% confidence level		$49^2 < 53^2 < 56^2$ [1a] $65^2 < 71^2 < 78^2$ [1d]	$146^2 < 208^2 < 373^2$ [1c] $121^2 < 195^2 < 572^2$ [1f]	$83^2 < 92^2 < 121^2$ [1b] $104^2 < 129^2 < 397^2$ [1e]	$36^2 < 39^2 < 42^2$ [1a] $111^2 < 122^2 < 157^2$ [1d]	The average variance and its 95% confidence interval is $91 < 122 < 274$
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$		55.8 [1a] 71.5 [1d]	227.7 [1c] 309.8 [1f]	91.6 [1b] 128.8 [1e]	47.8 [1a] 63.1 [1d] 120.7 [1g]	The average RMSE and its range is $56 < 124 < 310$

TABLE 10-4 (Concluded)

(24-Hour Average SO_2 $\mu g/m^3$)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location) Time and Location	Time Only	Location Only	Unmatched	Aggregate Statement of Model Performance and Accuracy
Associative-correlative	Pearson moment correlation coefficient r with Z statistic for the 95% confidence internal	307 [1a] 8 [1b] 47 [1c] 153 [1d] 4 [1e] 23 [1f] 31[1g]	0.66 < 0.72 < 0.77 [1a] 0.44 < 0.53 < 0.61 [1e]	0.62 < 0.77 < 0.87 [1c] 0.03 < 0.44 < 0.72 [1f]	-0.12 < 0.64 < 0.93 [1b] -0.95 < 0.13 < 0.97 [1e]		There is generally a significant positive correlation between predictions and observations at the 95% confidence level. The average correlation and its 95% confidence level is 0.11 < 0.54 < 0.81
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval						
Residual ($C_{obs} - C_{pred}$) differences degree of (bias)	Average residual with confidence interval of $\pm t_{\alpha}(S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		3.1 ± 1.0 [1a] 7.9 ± 1.1 [1d]	17.8 ± 6.5[1c] 21.3 ± 8.3[1f]	11.3 ± 2.6[1b] 15.7 ± 2.3[1e]	2.2 ± 1.0[1a] 5.8 ± 1.0[1d] 12.9 ± 1.3[1g]	There is significant positive bias at the 95% confidence level indicating a model underprediction; the average bias and its 95% confidence interval is 10.8 ± 2.7
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residuals as an estimator of σ_d^2 confidence interval $\nu e_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu e_d^2/\chi^2_{\nu,1-\alpha}$ where α = 95% confidence level		46 < 52 < 58 [1a] 37 < 43 < 52 [1d]	19 < 38 < 123 [1c] 16 < 42 < 357 [1f]	38 < 52 < 75 [1b] 27 < 40 < 70 [1e]	22 < 25 < 28 [1a] 18 < 21 < 25 [1d] 8.9 < 13.2 < 19 [1g]	The average variance of the residuals and its 95% con-confidence level is 20 < 38 < 133
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$ or average absolute error $\overline{|d|}$		7.8 [1a] 10.2 [1d]	18.7 [1c] 22.0 [1f]	13.4 [1b] 16.9 [1e]	5.4 [1a] 7.4 [1d] 13.4 [1g]	The average root-mean-square error and its range is 5.4 < 12.8 < 22.0

N.B. Numbers in brackets refer to citations in table 10.3.

TABLE 10-5. STATISTICAL MEASURES AND AGGREGATE STATEMENTS FOR RADM

(1-Hour Average SO_4 μg/m^3)

Performance Measure	General Statistical Measure	Sample Size (n)	Type of Match between Predictions and Observations (given in highest n observations as functions of time and location): Time and Location	Time Only	Location Only	Unmatched	Aggregate Statement of Model Performance and Accuracy
Associative-correlative	Pearson moment correlation coefficient r with Z statistic for the 95% confidence internal	13-19 [2] 175 [3]	$0.24 < 0.70 < 0.90$ [2] $0.36 < 0.70 < 0.88$ [3]				There is a significant positive correlation between predictions and observations at the 95% confidence level. The average correlation and 95% level is $0.30 < 0.76 < 0.89$
Frequency distribution difference	Goodness-of-fit--chi square or Kolmogorov-Smirnov statistic tests for the 95% confidence interval						
Residual ($C_{obs} - C_{pred}$) differences (bias)	Average residual with confidence interval of $\pm t_{\alpha} (S_o^2/N + S_p^2/N)^{1/2}$ where α = 95% confidence level		$-0.35 \pm ?$ [2] 2.9 ± 5.9 [3]				The bias for SO_2 is not significant at the 95% confidence level.
Residual ($C_{obs} - C_{pred}$) differences (noise)	Variance of residuals as an estimator of σ_d^2 confidence interval $\nu s_d^2/\chi^2_{\nu,\alpha} < \sigma_d^2 < \nu s_d^2/\chi^2_{\nu,1-\alpha}$, where α = 95% confidence level		$346 < 409 < 493$ [2]				None (see the statistic).
Residual ($C_{obs} - C_{pred}$) differences (gross error)	Root-mean-square error $RMSE_d$		2.76 [2] 20.4 [3]				None (see the statistic).

N.B. Numbers in brackets refer to the citations presented in table 10.3.

11 SUMMARY

The aggregate statements made in this study indicate average model performance tendencies for certain modeling attributes. The statements contain statistics that have been arithmetically averaged over several studies and over several prediction-observation/location-time matches. Care has been taken to extract statistics that are

> Representative of the whole evaluation study (e.g. single receptor statistics are ignored if there are many receptors that are not mentioned individually).

> The result of a modeling application that is a variant of the model or its intended usage.

In this survey of model performance evaluation studies, it was found that of the 42 models included, only 27 had one or more evaluation studies. The most evaluated models were

> CRSTER with 6 studies

> RAM with 7 studies

> AIRSHED with 6 studies

> HIWAY-1 with 5 studies.

Many of the model evaluation studies were for specific locations and dispersive conditions. Consequently, the aggregate statements represent limited statements about possible model performance and accuracy.

In general, model-to-model comparisons should be done with caution unless the models share several evaluation studies as the visibility models do. Nevertheless, it should be noted that many of the models whose evaluation studies are reported here are variations of the basic Gaussian plume model. In some cases, two such models may be virtually identical. Thus, performance evaluation results for one model may apply to others as well. As an example, six evaluation sutdies are cited for CRSTER while none are cited for MPTER. Since these two models are essentially identical (provided the appropriate options are selected), the CRSTER studies do, in fact, apply to MPTER. Other models included in this study may also be quite similar to each other, if not identical, so that evaluation studies for one often provide at least some information on the performance of another.

In many of the evaluation studies referenced, it was found that a substantial number of the non-AMS recommended statistics cannot be recast into AMS-recommended statistics. Nevertheless, several general observations can be made from the aggregate statements presented in this survey:

> Most of the statistical measures were computed from observations and predictions that are matched by time and location.

> Confidence intervals for bias were not reported very often and could not be calculated because averages and standard deviations of the sample set were not reported very often.

> Many evaluation studies are not clear in presenting or adequately defining the sample size.

> Cumulative frequency distributions are often limited to the upper 1, 5, or 10 percentiles. Thus, the sample size is too small to be statistically meaningful (e.g. < 10).

In some cases the evaluation study merely presents evaluation statistics and was not intended by the author(s) to be a comprehensive evaluation study but, rather, a limited application example. Consequently, care should be excercised when using a particular evaluation study to determine the performance and accuracy of a particular model.

REFERENCES

Abrams, I. B. (1979), Private Communication, 26 February 1979, to J. F. Bowers, H. E. Cramer Company, Inc.

Alabama Air Pollution Control Commission (1980), COMPTER Model User's Guide, Montgomery, Alabama.

Alabama Air Pollution Control Commission (1980), "Control Strategy Evaluation of COMPTER for the TVA Widows Creek Steam Plant," in Appendix C, Program Documentation for Multi-Source (SCSTER) Model by M. H. Malik and B. Baldwin, Southern Company Services, Inc., Atlanta, Gerogia.

Austin, D. I., A. W. Bealer, and W. R. Goodin (1981), "Random-Walk Advection and Dispersion Model (RADM) User's Manual," Dames and Moore, Los Angeles, California.

Benkley, C. W., and A. Bass (1979a), "Development of Mesoscale Air Quality Simulation Models. Vol. III. User's Guide to MESOPUFF (Mesoscale Puff) Model," EPA 600/7-80-058, Environmental Research and Technology, Inc., Concord, Massachusetts.

Benkley, C. W., and A. Bass (1979b), "Development of Mesoscale Air Quality Simulation Models. Vol II. User's Guide to Mesoplume (Mesoscale Plume Segment) Model," EPA 600/7-80-057, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Benson, P. E. (1980), "Background and Development of the CALINE3 Line Source and Dispersion Model," FHWA/CA/TL-80/31,.

Bensen, P. E. (1979), "CALINE3-A Versatile Dispersion Model for Predicting Air Pollutant Levels Near Highways and Arterial Streets," FHWA/CA/TL-79/23, Caltrans,.

Bergstrom, R. W., et al. (1981), "Comparison of the Observed and Predicted Visual Effects Caused by Power Plant Plumes," Atmos. Environ., Vol. 15, in press.

Bjorklund, J. R., and J. F. Bower, (1979), "User's Instructions for the Shortz and Longz Computer Programs (Vols. I and II)," TR-79-131-01, H. E. Cramer Co., University of Utah Research Park, Salt Lake City, Utah.

Bowers, J. F., and A. J. Anderson (1981), "An Evaluation Study for the Industrial Source Complex (ISC) Dispersion Model," EPA-450/4-81-002, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Bowers, J. F., J. R. Bjorklund, and C. S. Cheney (1979), "Industrial Source Complex (ISC) Dispersion Model User's Guide, Volumes I-II," EPA-450/4-79-031, Environmental Protection Agency,.

Bowers, J. F., Jr., and H. E. Cramer (1976), "Comparison of Calculated and Observed Characteristics of Plumes from Two Coal-Fired Power Plants Located in Complex Terrain," Third Symposium on Atmospheric Turbulence, Diffusion, and Air Quality, American Meteorological Society, Boston, Massachusetts.

Bowne, N. E., R. J. Londergan, D. H. Minott, and D. R. Murray (1981), "Interim Plume Model Validation Report," TRC Environmental Consultants, Inc., Wethersfield, Connecticut.

Busse, A. D., and J. R. Zimmerman (1973), User's Guide to the Climatological Dispersion Model, EPA-R4-73-024 (NTIS PB 227346/AS), Environmental Protection Agency, Addendum: EPA-450/3-77-015 (1977).

Calder, K. L. (1973), "A Climatological Model for Multiple Source Urban Air Pollution," in Appendix D, User's Guide for the Climatological Dispersion Model by A. D. Busse and J. R. Zimmerman, U. S. Environmental Protection Agency.

Chock, D. P. (1979), "Users Guide for the Simple Line-Source Model for Vehicle Exhaust Dispersion Near a Road," Environmental Science Department, General Motors Research Laboratories, Warren, Michigan.

Chock, D. P. (1978), "A Simple Line-Source Model for Dispersion near Roadways," Atmos. Environ., Vol. 12, pp. 823-829.

Chock, D. P. (1977), "General Motors Sulfate Dispersion Experiment: Assessment of the EPA-HIWAY Model," J. Air Poll. Control Assoc., Vol. 27, p. 39.

Corkum, D. A., and J. W. Bradstreet (1980), "A Sulfur Dioxide Modeling Comparison Study in the Rough Terrain of Northern New Hampshire near a Paper Mill," 73rd Annual Meeting, APCA, 22-27 June 1980, Montreal, Canada.

Dabberdt, W. F., F. L. Ludwig, and W. B. Johnson Jr. (1973), "Validation and Applications of an Urban Diffusion Model for Vehicular Pollutants," Atmos. Environ., Vol. 7, pp. 603-618.

Department of Transportation (1979), "Photochemical Modeling of Transportation Control Strategies, Vol. 1, Model Development, Performance Evaluation, and Strategy Assessment," Federal Highway Administration, Washington, D. C.

Drivas, P. J., et al. (1980), "ERT Visibility Model-Version 3-Technical Description and User's Guide," M-2020-001, Environmental Research and Technology, Inc., Concord, Massachusetts.

Egan, B. A., R. D'Errico, and C. Vaudo (1979), "Estimating Air Quality Levels in Regions of High Terrain Under Stable Atmospheric Conditions," Fourth Symposium on Turbulence, Diffusion, and Air Pollution, American Meteorological Society, 15-18 Jan. 1979, Reno Nevada.

Ellis, H. M., and P. C. Liu (1981), "Review of the Performance of the RAM Model in Predicting Highest Measured Concentrations," J. Air Poll. Control Assoc., Vol. 31, pp. 148-152.

Ellis, H., P. Liu, and G. Dalzell (1980), "Comparison Study of Measured and Predicted Concentrations with the RAM Model at Two Power Plants Along Lake Erie," Second Joint Conference on Applications of Air Pollution Meteorology, New Orleans, Louisiana.

Environmental Research and Technology, Inc. (1980a), "ERTAQ User's Guide," M-0186-001E, Environmental Research and Technology, Inc., Concord, Massachusetts.

Environmental Research and Technology, Inc. (1980b), "User's Guide to MPSDM Multiple Point Source Diffusion Model," M-0186-001-630, Environmental Research and Technology, Inc., Concord, Massachusetts.

Environmental Research and Technology, Inc. (1978), "Air Quality Model Validation Study for the Tampa Electric Big Bend Generating Station," P-2573, Environmental Research and Technology, Inc., Concord, Massachusetts.

Enviroplan, Inc. (1980), User's Manual For Enviroplan's Model 3141 and Model 4141, Enviroplan, Inc., West Orange, New Jersey.

Fabrik, A., R. Skarew, and J. Wilson (1977), "Point Source Model Evaluation and Development Study," A5-058-87, Science Applications, Inc., Westlake Village, California.

Fox, D. G. (1981), "Judging Air Quality Model Performance," Bull. Am. Meteorol. Soc., Vol. 62, pp. 599-609.

Godden, D., E. Luvmann, A. Lloyd, B. Nitta, and C. Kwah (1980), "Further Evaluation of ELSTAR (Environmental Lagrangian Simulator of Transport and Atmospheric Reactions) in Los Angeles under Different Meteorological Conditions," Environmental Research and Technology, Inc., Westlake Village, California.

Goodin, W. R., et al. (1980), "Evaluation and Application of RADM, a Lagrangian, Random-Walk Model for Advection and Dispersion," Dames and Moore, Los Angeles, California.

Guldberg, P. H., and C. W. Kern (1978), "A Comparative Validation of the RAM and PTMTP Models for Short-Term Concentrations in Two Urban Areas," J. Air Poll. Control Assoc., Vol. 28, pp. 907-910.

Hales, J. M., and S. L. Sutter (1973), "Solubility of Sulfur Dioxide in Water at Low Concentrations," Atmos. Environ., Vol. 7, pp 997-1105.

Hilst, G. R. (1978), "Plume Model Validation," EPRI EA-917-5y, Electric Power Research Institute, Palo Alto, California.

Hodanbosi, R. G., and L. K. Peters (1981), "Evaluation of RAM Model for Cleveland, Ohio," J. Air Poll. Control Assoc., Vol. 31, pp. 253-255.

Johnson, C. E., et al. (1980), "User's Manual for the Plume Visibility Model (PLUVUE)," EPA 450/4-80-032, final report, Systems Applications, Inc., San Rafael, California.

Johnson, W. B., F. L. Ludwig, W. F. Dabberdt, and R. J. Allen (1973), "An Urban Diffusion Simulation Model for Carbon Monoxide," J. Air Poll. Control Assoc., Vol. 977), "Point Source Model Evaluation and Development Study," A5-058-87, Science Applications, Inc., Westlake Village, California.

Klug, W. (1974), "Dispersion from Tall Stacks," 5th NATO/CCMS International Technical Meeting on Air Pollution Modeling, Denmark.

Koch, R. C., and G. E. Fisher (1973), "Evaluation of the Multiple Source Gaussian Plume Diffusion Model-Phase I," EPA-650/4-75-018a, Environmental Protection Agency.

Kummler, R. H., et al. (1979), "A Comparative Validation of the RAM and Modified SAI Models for Short-Term SO_2 Concentrations in Detroit," J. Air Poll. Control Assoc., Vol. 29, pp. 720-723.

Lindgren, B. W. (1976), Statistical Theory, 3d ed. (Macmillan Publishing Co. Inc., New York, New York).

Liu, M. K. (1981), Private Communication, Systems Applications, Inc., San Rafael, California.

Londergan, R. J., et al. (1980), "An Evaluation of Short-Term Air Quality Models Using Tracer Study Data," 4333, American Petroleum Institute, Washington, D.C.

Lloyd, A. C., et al. (1979), "Development of the ELSTAR Photochemical Air Quality Simulation Model and Its Evaluation Relative to the LARPP Data Base," P-5287-500, Environmental Research and Technology, Inc., Westlake Village, California.

Maldonado, C., and J. A. Bullin (1977), "Modeling of Carbon Monoxide Dispersion from Roads," Environ. Sci. Technol., Vol. II, p. 1071.

Malik, M. H., and B. Baldwin (1980), "Program Documentation for Multi-Source (SCSTER) Model," EN7408SS, Southern Company Services, Inc., Atlanta, Georgia.

McNidar, R. T. (1977), "Variability Analysis of Long-term Dispersion Models," Joint Conf. on Applications of Air Pollution Meteorology, American Meteorology Society, 29 Nov.-2 Dec., 1977, Salt Lake City, Utah.

Mills, M. T., R. Caiazza, D. D. Hergert, and D. A. Lynn, (1980) "Evaluation of Point Source Dispersion Models," Teknekron Research, Inc., Waltham, Massachusetts.

Mills, M., and F. A. Record (1975), "Comprehensive Analysis of Time-Concentration Relationships and the Validation of a Single Source Dispersion Model," EPA-450/3-75-083, Environmental Protection Agency, Washington, D. C.

Mills, M., and R. Stern (1975), "Model Validation and Time-Concentration Analysis of Three Power Plants," EPA-450/3-76-002, Environmental Protection Agency.

Morgenstern, P., M. J. Geraghty, and A. McKnight (1979), "A Comparative Study of the RAM (Urban) and RAMR (Rural) Models for Short-term SO_2 Concentrations in Metropolitan Indianapolis," 79-2.2, 72nd Annual Meeting of the Air Pollution Control Association, Cincinnati, Ohio.

Morris, C. S. (1980), "User's Guide for RTDM.WC," M-0186-000R, Environmental Research and Technology, Inc., Concord, Massachusetts.

Morris, C. S., C. W. Benkley, and A. Bass (1979), "Development of Mesoscale Air Quality Simulation Models: Volume IV. User's Guide to Mesogrid (Mesoscale Grid) Model," EPA-600/7-80-059, U.S. Environmental Protection Agency, Washington, D. C.

Moser, J. H. (1979), *MULTIMAX: An Air Dispersion Modeling Program for Multiple Sources, Receptors and Concentration Averages*, Shell Development Co., Houston, Texas.

New Mexico Environmental Improvement Division, (1981), *Control Strategy Demonstration: New Mexico Air Quality Control Regulation 602, Coal Burning Equipment-Sulfur Dioxide*, (New Mexico Environmental Improvement Division, Santa Fe, New Mexico).

Nieuwstadt, F. T. M., H. Van Dop (1975), "Validation of a Multiple Source Diffusion Model Based on the Gaussian Plume Model," 6th NATO/CCMS Int. Technical Meeting on Air Pollution Modeling, Frankfurt, West Germany.

Noll, K. E., et al. (1978), "A Comparison of Three Highway Line Source Dispersion Model," *Atmos. Environ.*, Vol 12, pp. 1323-1329.

Norton, C. L., and A. Roffman (1979), "User's Guide for the Particulate Matter and Sulfur Dioxide Diffusion-Deposition Model (PASDIM)," Energy Impact Associates, Pittsburg, Pennsylvania.

Pacific Gas and Electric Co. (1981), *User's Manual for Pacific Gas and Electric PLUME5 Model*, Pacific Gas and Electric Co., San Francisco, California.

Petersen, W. B., (1980), "User's Guide for HIWAY-2: A Highway Air Pollution Model," EPA-600/8-80-018, Environmental Protection Agency.

Petersen, W. B. (1978), "User's Guide for PAL: A Gaussian-Plume Algorithm for Point, Area, and Line Source," EPA-600/4-78-013, Environmental Protection Agency.

Pierce, T. E., and D. B. Turner (1980), "User's Guide for MPTER," EPA-600/8-80-016, Environmental Protection Agency, Washington D. C.

Ragland, K. W., and K. E. Wilkening (1980), "Intermediate-Range Grid Model for Atmospheric Sulfur Dioxide and Sulfate Concentrations and Depositions," CR806898-01-5, Environmental Research Laboratory, U. S. Environmental Protection Agency, Duluth, Minnesota.

Rao, S. T., and J. R. Visalli (1981), "On the Comparative Assessment of the Performance of Air Quality Models," J. Air Poll. Control Assoc., Vol. 31, pp. 851-860.

Rao, S. T., G. Sistla, M. T. Keenan, and J. S. Wilson (1980), "An Evaluation of Some Commonly Used Highway Dispersion Models," J. Air Poll. Control Assoc., Vol. 30, p.239.

Rao, S. T., and M. T. Keenan (1980), "Suggestions for Improvement of the EPA-Highway Model," J. Air Poll. Control Assoc., Vol. 30, pp. 60-69.

Reid, L. E., S. D. Reynolds, and D. A. Latimer (1980), "Evaluation of Airshed Model Performance in Tulsa," SAI No. 223-ES80-164, Systems Applications, Inc., San Rafael, California.

Reynolds, S. D., T. W. Tesche, and L. E. Reid (1979), "An Introduction to the SAI Airshed Model and Its Usage," SAI-EF79-31, Systems Applications, Inc., San Rafael, California.

Roffman, A. (1979), "A Diffusion Study of Particulate Matter and SO_2 in Allegheny County, Pennsylvania," 72nd Annual Meeting of the Air Pollution Control Association, Cincinnati, Ohio.

Roth, P. M., et al. (1981), "Evaluation of the Impacts of NO_x Emissions from the Proposed Montezuma Generating Stations on Oxidant Concentrations in an Area Between Lodi and Stockton," SAI-80R2-EF81-65, Systems Applications, Inc., San Rafael, California.

Ruff, R. E. (1980), "Evaluation of the RAM Using the RAPS Data Base," Contract 68-02-2770, SRI International, Menlo Park, California.

Runchal, A. K. (1980), "A Random Walk Atmospheric Dispersion Model for Complex Terrain and Meteorological Conditions," Second Joint Conference on Applications of Air Pollution Meteorology and Second Conference on Industrial Meteorology, 28 March 1980, New Orleans, Louisiana.

Schulman, L. L., and J. S. Scire (1980), "Development of an Air Quality Dispersion Model for Aluminum Reduction Plants," P-7304A, Environmental Research and Technology, Inc., Concord, Massachusetts.

Schulman, L. L., and J. S. Scire (1980), "Buoyant Line and Point Source (BLP) Dispersion Model User's Guide,", P-7304B, Environmental Research and Technology, Inc., Concord, Massachusetts.

Scire, J. S., and L. L. Schulman (1981), "Evaluation of the BLP and ISC Models with SF_6 Tracer Data and SO_2 Measurements at Aluminum Reduction Plants," ARCA Specialty Conference on Dispersion Modeling for Complex Sources .

Scott, B. C. (1978), "Parameterization of Sulfate Removal by Precipitation," J. of Appl. Meteor., Vol. 17, pp 1375-1389.

Sklarew, R. C., and V. Mirabella (1979), "Experience in Impact Modeling of Complex Terrain," Fourth Symposium on Turbulence, Diffusion, and Air Pollution, American Meteorological Society, Boston, Massachusetts.

Sklarew, R. C., and K. T. Tran (1978), "Cobb Valley Rough Terrain Modeling Results," Geothermal Seminar-78, Sacramento, California.

Stewart, D. A. (1981), Private Communication, Systems Applications, Inc., San Rafael, California.

Stewart, D. A., M. K. Liu, and D. Henderson (1981), "An Assessment of Relative Contribution of Wet and Dry Deposition via a Long-Range Transport Model," Fifth Symposium on Turbulence, Diffusion, and Air Pollution, 9-13 March 1981, Atlanta, Georgia.

Stewart, D. A., and M. K. Liu (1981), "Development and Application of a Reactive Plume Model," Atmos. Environ., in press

Systems Applications, Inc. (1981), Impacts of NO_x Control upon Ozone Concentrations in Three Environments," Report to the Electric Power Research Institute, Systems Applications, Inc., San Rafael, California.

TRW Systems Group (1969), "Air Quality Display Model," NTIS #PB189194 .

Tesche, T. W., et al. (1981), "The Sensitivity of Complex Photochemical Model Estimates to Detail in Input Information," SAI No. 330R-EF81-5, EPA No. 68-02-2870, Systems Applications, Inc. San Rafael, California.

Tesche, T. W., and C. S. Burton (1978), "Simulated Impact of Alternative Emissions Control Strategies on Photochemical Oxidants in Los Angeles Vol. I--Preliminary Discussion of Results," SAI-EF-78-22r, Systems Applications, Inc., San Rafael, California.

Texas Air Control Board (1979), User's Guide to The Texas Episodic Model, (Texas Air Pollution Control Board, Austin, Texas).

Tran, K. T., and R. C. Sklarew (1979), "User Guide to IMPACT: An Integrated Model for Plumes and Atmospheric Chemistry in Complex Terrain," Form and Substance, Inc., Westlake Village, California.

Turner, D. B. (1980), "A Post-Screening Terrain Model," Chicago Workshop on Air Quality Models, Chicago, Illinois.

Turner, D. B., J. R. Zimmerman, and A. D. Busse (1973), "An Evaluation of Some Climatological Dispersion Models," in Appendix E of User's Guide to the Climatological Dispersion Model, Environmental Protection Agency, Research Triangle Park, North Carolina.

Turner, D. B., and J. H. Novak (1978), "User's Guide for RAM," EPA-600/8-78-016a, Environmental Protection Agency,.

U. S. Environmental Protection Agency (1977), User's Manual for Single-Source (CRSTER) Model, EPA-450/2-77-013 (Environmental Protection Agency).

Wang, I. T., T. L. Waldron, and B. R. Bushey (1980), "A Mesoscale Transport and Dispersion Model for Industrial Plumes," Second Point Conference on Applications of Air Pollution Meteorology and Second Conference on Industrial Meteorology, 28 March 1980, New Orleans, Louisiana.

Zimmerman, J. R., and R. S. Thompson (1975), "Users Guide for HIWAY, A Highway Air Pollution Model," EPA-650/4-74-008, Environmental Protection Agency .